ELECTROMAGNETISM

0 Vector Analysis

0.1 INTRODUCTION

The study of the theory of electromagnetic fields needs the knowledge and the usage of certain mathematical concepts as tools for better understanding of the subject. Elements of differential and integral calculus are essential for defining the basic concepts of the theory. Some other mathematical background such as vector algebra and calculus, complex numbers, conformal transformation, some transforms, matrix algebra, and aspects of numerical analysis also help very significantly in achieving clarity of understanding—thereby making possible a simple formal development of the subject. In the present study, the emphasis is on the understanding of the physical aspects of the theory rather than the lengthy mathematical manipulations. In such an approach, the concepts of vectors play a central role in the study of electromagnetic fields.

In the present text, the vectors have been used as mathematical tools from the very beginning. Though in today's pre-engineering mathematics courses the students are introduced to vector algebra, it is felt that a concise introduction here to the algebra and the calculus of vectors would greatly help and simplify the study of the subject as presented in this book. The present introduction to vector methods will be utilized to emphasize the physics of, and its application to, practical systems and situations.

0.2 VECTORS AND VECTOR ALGEBRA

There are two classes of physical quantities, each with its characteristic properties and an appropriate algebra.

Scalar quantities. These have magnitude only and do not involve direction. To specify a scalar quantity completely, it is necessary to know (i) a unit of the same kind and (ii) a number stating how many times the unit is contained in the quantity, i.e. a scalar can be specified by a single number. Examples of scalar quantities are: mass, volume, density, temperature, electric potential, charge, and so forth. They are indicated by non-bold face letters, for instance, S.

Vector quantities. These have both magnitude and direction. For the complete specification of a vector quantity, it is necessary to know (i) a unit of the same kind, (ii) a number giving the magnitude of the quantity in terms of this unit, and (iii) a statement of the direction. Examples of vector quantities are: velocity, acceleration, force, stress, displacement, electric force, magnetic induction, and so on. They are indicated by boldface type, for instance, $\mathbf{V}$.

The product of a vector and a scalar is defined as a vector whose magnitude is equal to the numerical product of the magnitude of the original vector and the magnitude of the scalar, and whose direction coincides with that of the original vector or is opposite to it depending on whether the scalar is positive or negative. Thus a vector $\times$ (–1) is equivalent to reversing its direction, i.e.

$$\mathbf{V} \times (-1) = -\mathbf{V} \tag{0.1}$$

There are two classes of quantities represented by vectors:

(i) Force, displacement, velocity, and the like in which the vector is drawn in the direction of the quantity concerned. That is, linear action in a particular direction. Such vectors are called *polar vectors*.

(ii) Angular velocity, angular acceleration, and the like in which rotary action of some kind takes place about an axis (the positive direction of the vector is obtained by the rule of the right-handed screw). These are called *axial vectors*. Note that the vectors cannot represent the finite rotations of a body about a fixed axis.

There is little difference in the mathematical treatment of these two classes of vectors.

0.2.1 Addition and Subtraction of Vectors

The vector addition follows the parallelogram law, i.e.

$$\mathbf{V} = \mathbf{A} + \mathbf{B} \tag{0.2}$$

where $$|\mathbf{V}| = \sqrt{A^2 + B^2 - 2AB\cos\theta}$$

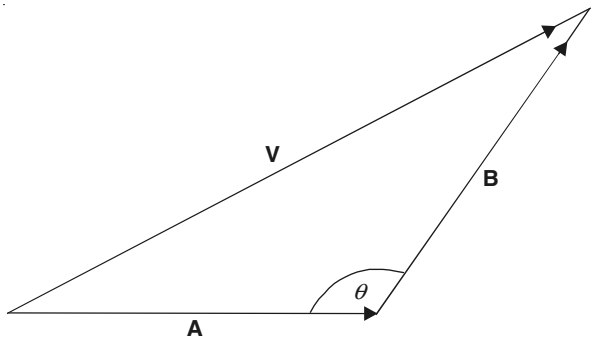


Figure 0.1 Vector addition.

Vector addition is commutative, i.e.

$$\mathbf{V} = \mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$$

Vector addition is also associative, i.e.

$$(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{C}) + \mathbf{B}, \text{ and so on.}$$

Vector subtraction is obtained by reversing the sign of the vector that is being subtracted.

0.2.2 Components of a Vector

By reversing the process of addition, any vector can be expressed as a sum of n vectors, of which $(n - 1)$ are arbitrary and the last one closes the polygon. In general, the vectors need not be coplanar.

An useful example is the component vectors along the orthogonal axes of the Cartesian coordinate system. This is because a vector in a three-dimensional space is completely known if its origin is specified and its projections on three mutually perpendicular directions are known. Hence a given vector can be expressed in terms of its three unit vectors in the directions of an orthogonal coordinate system, i.e.

$$\mathbf{V} = \mathbf{V}_x + \mathbf{V}_y + \mathbf{V}_z = \mathbf{i}_x V_x + \mathbf{i}_y V_y + \mathbf{i}_z V_z \tag{0.3}$$

where $\mathbf{i}_x$, $\mathbf{i}_y$, $\mathbf{i}_z$ are the unit vectors in the directions x, y, and z, respectively. In the present discussion, we will follow the right-handed screw convention for the positive directions of x, y, and z. (The need for a consistent coordinate system will become obvious when we consider vector multiplications as the vectors are directional quantities.)

From Figure 0.2,

$$\begin{aligned} V_x &= V \cos \theta_x \\ V_y &= V \cos \theta_y \\ V_z &= V \cos \theta_z \end{aligned} \tag{0.4}$$

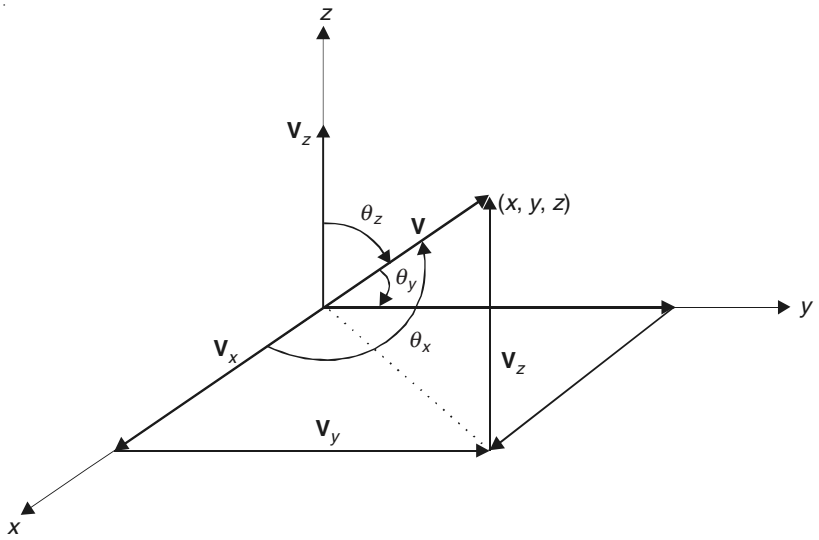


Figure 0.2 Orhogonal components of a vector.

0.2.3 Multiplication of Vectors

There are two types of vector multiplication.

0.2.3.1 The scalar product of vectors

The *scalar product* of the vectors **A** and **B** is defined as

$$\mathbf{A} \cdot \mathbf{B} = AB \cos \theta \tag{0.5}$$

where θ is the angle between the vectors **A** and **B** (Figure 0.3). Thus $\mathbf{A} \cdot \mathbf{B}$ (also called the *dot product*) is a scalar and is the numerical product of the magnitude of **B** and the component of **A** in the direction of **B**, or is the product of the magnitude of **A** and the component of **B** in the direction of **A**.

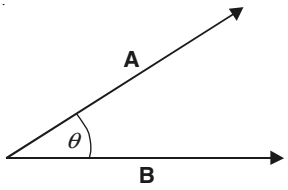


Figure 0.3 Scalar product of the vectors **A** and **B**.

A common physical example in mechanics is that of work done by a force **A** in moving an object through a distance **B** along a straight line. Thus,

Work done = magnitude of the displacement × the component of the force in the direction of the displacement.

Note that:

$$\mathbf{A} \cdot \mathbf{B} = \mathbf{B} \cdot \mathbf{A} \tag{0.6}$$

That is, the dot, or scalar, multiplication obeys the commutative law.

$$(\mathbf{A} + \mathbf{B}) \cdot \mathbf{C} = \mathbf{A} \cdot \mathbf{C} + \mathbf{B} \cdot \mathbf{C} \tag{0.7}$$

That is, the dot product obeys the distributive law.

$$(\mathbf{A} \cdot \mathbf{B})\mathbf{C} \neq \mathbf{A}(\mathbf{B} \cdot \mathbf{C}) \tag{0.8}$$

That is, the dot product does not obey the associative law.

Also, applying the dot product to the unit vectors of the Cartesian coordinate system, we get

$$\mathbf{i}_x \cdot \mathbf{i}_x = \mathbf{i}_y \cdot \mathbf{i}_y = \mathbf{i}_z \cdot \mathbf{i}_z = 1 \tag{0.9}$$

and

$$\mathbf{i}_x \cdot \mathbf{i}_y = \mathbf{i}_y \cdot \mathbf{i}_z = \mathbf{i}_z \cdot \mathbf{i}_x = 0 \tag{0.10}$$

Example 0.1 Evaluate $\mathbf{A} \cdot \mathbf{B}$ in terms of the components of **A** and **B** in the coordinate system.

SOLUTION

$$\mathbf{A} \cdot \mathbf{B} = (\mathbf{i}_x A_x + \mathbf{i}_y A_y + \mathbf{i}_z A_z) \cdot (\mathbf{i}_x B_x + \mathbf{i}_y B_y + \mathbf{i}_z B_z) = A_x B_x + A_y B_y + A_z B_z \tag{0.11}$$

0.2.3.2 The vector product of vectors

The *vector product*, or cross product, of **A** and **B** is a vector **C**, i.e.

$$\mathbf{C} = \mathbf{A} \times \mathbf{B} \tag{0.12}$$

such that the vector **C** is normal to the plane defined by **A** and **B** (Figure 0.4). The positive direction of **C** is defined by the right-handed screw rule, that is, as the screw rotates from **A** to **B** through the smaller angle θ, the motion of the screw is in the positive direction of **C**.

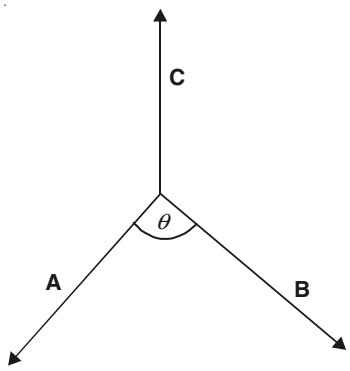


Figure 0.4 Vector product of the vectors **A** and **B**.

The magnitude of **C** is defined as

$$C = AB \sin \theta \tag{0.13}$$

Thus C is equal to the area of the parallelogram formed by **A** and **B**. It also follows from the definition that

$$\mathbf{A} \times \mathbf{B} = -\mathbf{B} \times \mathbf{A} \tag{0.14}$$

Thus the vector product does not obey the commutative law. Also,

$$(\mathbf{A} \times \mathbf{B}) \times \mathbf{C} \neq \mathbf{A} \times (\mathbf{B} \times \mathbf{C}) \tag{0.15}$$

Therefore, the vector product does not obey the associative law as well.

A diagrammatic proof. Let the vector **A** be parallel to the xy-plane, the vector **B** be parallel to the y-axis, and the vector **C** be parallel to the yz-plane [see Figure 0.5(a)]. In this case, $(\mathbf{A} \times \mathbf{B})$ is parallel to the z-axis and so the product $(\mathbf{A} \times \mathbf{B}) \times \mathbf{C}$ will be parallel to the x-axis. Now $(\mathbf{B} \times \mathbf{C})$ is parallel to the x-axis, and so $\mathbf{A} \times (\mathbf{B} \times \mathbf{C})$ will be parallel to the z-axis [see Figure 0.5(b)].

$$\therefore \qquad (\mathbf{A} \times \mathbf{B}) \times \mathbf{C} \neq \mathbf{A} \times (\mathbf{B} \times \mathbf{C})$$

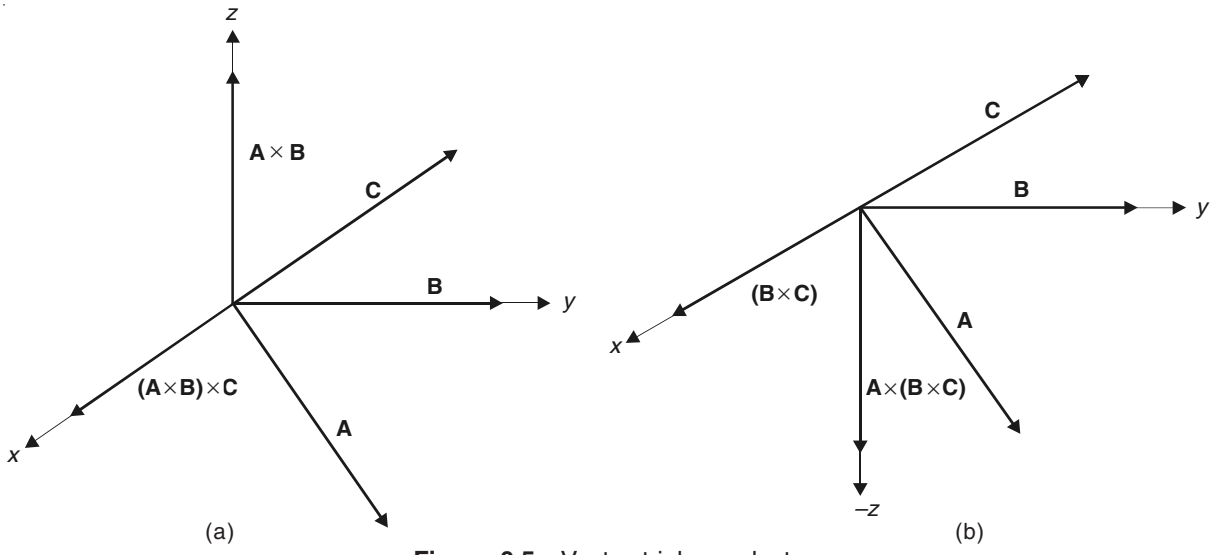


Figure 0.5 Vector triple product.

But the vector product obeys the distributive law, i.e.

$$(\mathbf{A} + \mathbf{B}) \times \mathbf{C} = \mathbf{A} \times \mathbf{C} + \mathbf{B} \times \mathbf{C} \tag{0.16}$$

Applying vector multiplication to the unit vectors of the rectangular Cartesian coordinate axes, we get

$$\mathbf{i}_x \times \mathbf{i}_x = \mathbf{i}_y \times \mathbf{i}_y = \mathbf{i}_z \times \mathbf{i}_z = 0 \tag{0.17}$$

and

$$\mathbf{i}_x \times \mathbf{i}_y = \mathbf{i}_z, \qquad \mathbf{i}_y \times \mathbf{i}_z = \mathbf{i}_x, \qquad \mathbf{i}_z \times \mathbf{i}_x = \mathbf{i}_y \tag{0.18}$$

Using these unit vectors, the cross product of the vectors **A** and **B** can be expressed as

$$\begin{aligned}\mathbf{A} \times \mathbf{B} &= (\mathbf{i}_x A_x + \mathbf{i}_y A_y + \mathbf{i}_z A_z) \times (\mathbf{i}_x B_x + \mathbf{i}_y B_y + \mathbf{i}_z B_z) \\ &= \mathbf{i}_x(A_y B_z - A_z B_y) + \mathbf{i}_y(A_z B_x - A_x B_z) + \mathbf{i}_z(A_x B_y - A_y B_x) \\ &= \begin{bmatrix} \mathbf{i}_x & \mathbf{i}_y & \mathbf{i}_z \\ A_x & A_y & A_z \\ B_x & B_y & B_z \end{bmatrix}\end{aligned} \tag{0.19}$$

0.2.3.3 Triple products

The triple scalar product. Given three vectors **A**, **B**, and **C**, the product $\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C})$ is called the triple scalar product. Then

$$\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = \begin{bmatrix} A_x & A_y & A_z \\ B_x & B_y & B_z \\ C_x & C_y & C_z \end{bmatrix} \tag{0.20}$$

This is the volume of the parallelepiped formed by the three vectors. Also, for any three vectors **A**, **B**, and **C**

$$\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = \mathbf{B} \cdot (\mathbf{C} \times \mathbf{A}) = \mathbf{C} \cdot (\mathbf{A} \times \mathbf{B}) = -\mathbf{A} \cdot (\mathbf{C} \times \mathbf{B}) = -\mathbf{B} \cdot (\mathbf{A} \times \mathbf{C}) = -\mathbf{C} \cdot (\mathbf{B} \times \mathbf{A}) \tag{0.21}$$

Since, when the three vectors are coplanar, the volume of the parallelepiped is zero; this scalar product therefore also defines the condition required for coplanar vectors.

The triple vector product. The product $\mathbf{A} \times (\mathbf{B} \times \mathbf{C})$

$$\begin{aligned}&= (\mathbf{i}_x A_x + \mathbf{i}_y A_y + \mathbf{i}_z A_z) \times \{\mathbf{i}_x(B_y C_z - B_z C_y) + \mathbf{i}_y(B_z C_x - B_x C_z) + \mathbf{i}_z(B_x C_y - B_y C_x)\} \\ &= \mathbf{i}_x\{A_y(B_x C_y - B_y C_x) - A_z(B_z C_x - B_x C_z)\} + \mathbf{i}_y\{A_z(B_y C_z - B_z C_y) - A_x(B_x C_y - B_y C_x)\} + \\ &\quad \mathbf{i}_z\{A_x(B_z C_x - B_x C_z) - A_y(B_y C_z - B_z C_y)\} \\ &= \mathbf{i}_x\{B_x(A_y C_y + A_z C_z) - C_x(A_y B_y + A_z B_z) + B_x A_x C_x - C_x A_x B_x\} + \\ &\quad \mathbf{i}_y\{B_y(A_z C_z + A_x C_x) - C_y(A_z B_z + A_x B_x) + B_y A_y C_y - C_y A_y B_y\} + \\ &\quad \mathbf{i}_z\{B_z(A_x C_x + A_y C_y) - C_z(A_x B_x + A_y B_y) + B_z A_z C_z - C_z A_z B_z\} \\ &= \mathbf{i}_x\{(\mathbf{A} \cdot \mathbf{C})B_x - (\mathbf{A} \cdot \mathbf{B})C_x\} + \mathbf{i}_y\{(\mathbf{A} \cdot \mathbf{C})B_y - (\mathbf{A} \cdot \mathbf{B})C_y\} + \mathbf{i}_z\{(\mathbf{A} \cdot \mathbf{C})B_z - (\mathbf{A} \cdot \mathbf{B})C_z\} \\ &= (\mathbf{A} \cdot \mathbf{C})\mathbf{B} - (\mathbf{A} \cdot \mathbf{B})\mathbf{C}\end{aligned}$$

$$\therefore \quad \mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = (\mathbf{A} \cdot \mathbf{C})\mathbf{B} - (\mathbf{A} \cdot \mathbf{B})\mathbf{C} \tag{0.22}$$

Similarly,

$$\begin{aligned}(\mathbf{A} \times \mathbf{B}) \times \mathbf{C} &= -\mathbf{C} \times (\mathbf{A} \times \mathbf{B}) \\ &= -\{(\mathbf{C} \cdot \mathbf{B})\mathbf{A} - (\mathbf{C} \cdot \mathbf{A})\mathbf{B}\} \\ &= (\mathbf{A} \cdot \mathbf{C})\mathbf{B} - (\mathbf{B} \cdot \mathbf{C})\mathbf{A} \end{aligned} \tag{0.23}$$

0.3 THREE ORTHOGONAL COORDINATE SYSTEMS

So far, we have used the rectangular Cartesian coordinate system to represent vectors. There are at least two more orthogonal systems, which are widely used for vector operations. However, it must be remembered that vectors are physically independent of any system of coordinates by which they may be expressed. They are said to be 'invariant' and the coordinate systems are really unnecessary in vector analysis. However since a vector in the three-dimensional space is completely represented by its three orthogonal components, its representation by using suitable orthogonal coordinate axes greatly helps in understanding as well as in making mathematical manipulations more concise and simple. Here we will briefly describe the basic features of the most widely used coordinate systems. For a relatively more detailed discussion of the generalized coordinate systems and the corresponding vector operations, the reader is advised to refer to Appendix 4.

0.3.1 Rectangular Cartesian Coordinates (x, y, z)

We shall use the right-handed Cartesian system, i.e. if x turns to y through the smaller angle between them, then the positive direction of z is given by the right-handed screw rule (see Figure 0.6 for right- and left-handed coordinate systems).

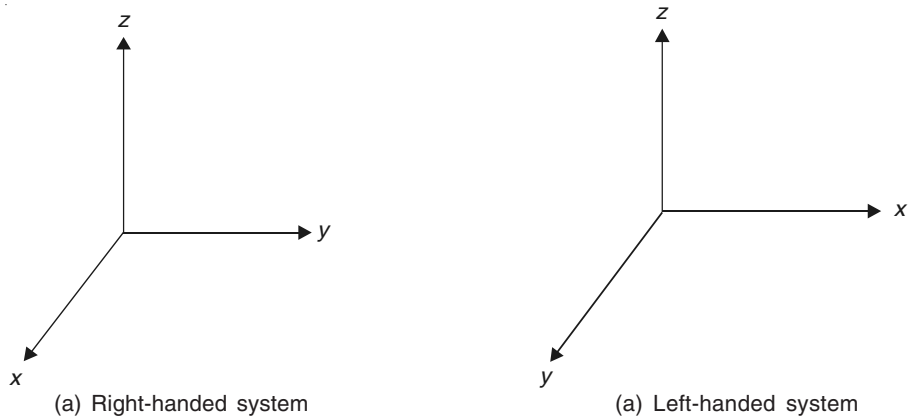


Figure 0.6 Right-handed and left-handed Cartesian coordinate systems.

The length elements are: dx, dy, dz (or δx, δy, δz).

The unit vectors in the directions of the coordinate axes are: $\mathbf{i}_x$, $\mathbf{i}_y$, $\mathbf{i}_z$ (or $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$).

The surface elements in the coordinate planes are: $\delta x \delta y$, $\delta y \delta z$, $\delta z \delta x$. The volume element is: $\delta x \delta y \delta z$.

0.3.2 Cylindrical Polar Coordinate System (r, ϕ, z)

This system is obtained by taking a radius vector in the xy-plane and rotating it about the z-axis, that is, translation of the xy-plane normal to the z-axis (Figure 0.7).

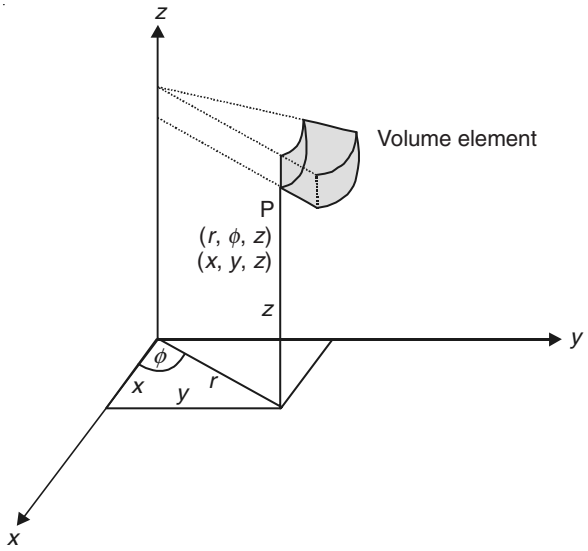


Figure 0.7 Elements in a cylindrical polar coordinate system.

The length elements are: δr, $r\delta\phi$, δz.
The surface elements parallel to the coordinate planes are: $r\delta r\delta\phi$, $r\delta\phi\delta z$, $\delta r\delta z$.
The volume element is: $r\delta r\delta\phi\delta z$.
The relationships between the Cartesian and the cylindrical coordinate systems are:

$$r = \sqrt{x^2 + y^2}$$
$$\phi = \tan^{-1}\frac{y}{x}$$
$$z = z$$

and

$$x = r\cos\phi$$
$$y = r\sin\phi$$
$$z = z \tag{0.24}$$

0.3.3 Spherical Polar Coordinate System (ρ, θ, ϕ)

This system is obtained by rotating the radius vector about the origin (Figure 0.8).
The length elements along the coordinate axes are: $\delta\rho$, $\rho\delta\theta$, $\rho\sin\theta\,\delta\phi$.
The unit vectors along the coordinate axes are: $\mathbf{i}_\rho$, $\mathbf{i}_\theta$, $\mathbf{i}_\phi$.
The surface elements parallel to the coordinate planes are: $\rho\delta\rho\delta\theta$, $\rho^2\sin\theta\delta\theta\delta\phi$, $\rho\sin\theta\delta\rho\delta\phi$.
The volume element is: $\rho^2\sin\theta\,\delta\rho\,\delta\theta\,\delta\phi$.
The relationships between the spherical and the Cartesian coordinate systems are:

$$\rho = \sqrt{x^2 + y^2 + z^2}$$

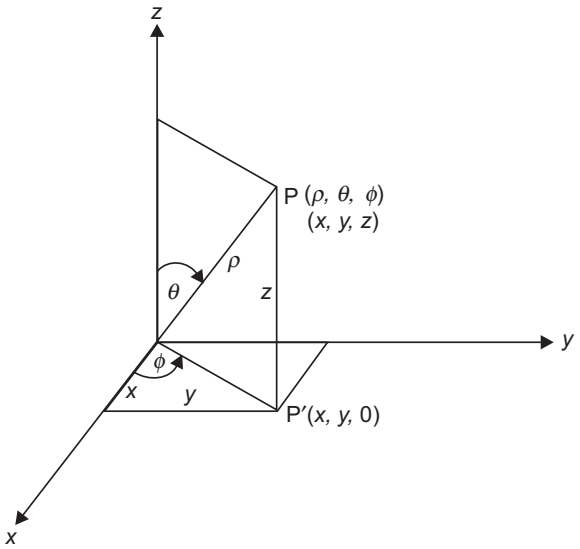


Figure 0.8 Spherical polar coordinate system.

$$\theta = \tan^{-1} \frac{\sqrt{x^2 + y^2}}{z}$$

$$\phi = \tan^{-1} \frac{y}{x}$$

and

$$\begin{aligned} x &= \rho \sin \theta \cos \phi \\ y &= \rho \sin \theta \sin \phi \\ z &= \rho \cos \theta \end{aligned} \tag{0.25}$$

The relationships between the spherical and the cylindrical coordinate systems are:

$$\rho = \sqrt{r^2 + z^2}$$

$$\theta = \tan^{-1} \frac{r}{z}$$

$$\phi = \phi$$

and

$$\begin{aligned} r &= \rho \sin \theta \\ \phi &= \phi \\ z &= \rho \cos \theta \end{aligned} \tag{0.26}$$

0.4 VECTOR CALCULUS

0.4.1 Differentiation of Vectors

Suppose that the vector $\mathbf{V}$ is a vector function of the scalar variable t. When t changes from t to $t + \delta t$, $\mathbf{V}$ becomes $\mathbf{V} + \delta\mathbf{V}$, and the average rate of change of $\mathbf{V}$ with respect to t as $\delta t \to 0$ is

$$\frac{d\mathbf{V}}{dt} = \lim_{\delta t \to 0}\left(\frac{\delta \mathbf{V}}{\delta t}\right) \tag{0.27}$$

In the Cartesian coordinates, $\mathbf{V} = \mathbf{i}_x V_x + \mathbf{i}_y V_y + \mathbf{i}_z V_z$, then

$$\frac{d\mathbf{V}}{dt} = \mathbf{i}_x\left(\frac{dV_x}{dt}\right) + \mathbf{i}_y\left(\frac{dV_y}{dt}\right) + \mathbf{i}_z\left(\frac{dV_z}{dt}\right) \tag{0.28}$$

For example, If $\mathbf{V}$ is the vector of displacement and t represents time, then $d\mathbf{V}/dt$ is the velocity and $d^2\mathbf{V}/dt^2$ is the acceleration.

If $\mathbf{V} = \mathbf{A} + \mathbf{B}$, then

$$\frac{d\mathbf{V}}{dt} = \frac{d\mathbf{A}}{dt} + \frac{d\mathbf{B}}{dt} \tag{0.29}$$

If the scalar $S = \mathbf{A} \cdot \mathbf{B}$ is differentiated, then

$$\frac{dS}{dt} = \frac{d\mathbf{A}}{dt} \cdot \mathbf{B} + \mathbf{A} \cdot \frac{d\mathbf{B}}{dt} \tag{0.30}$$

If the vector $\mathbf{V} = \mathbf{A} \times \mathbf{B}$ is differentiated, then

$$\frac{d\mathbf{V}}{dt} = \frac{d\mathbf{A}}{dt} \times \mathbf{B} + \mathbf{A} \times \frac{d\mathbf{B}}{dt} \tag{0.31}$$

This relation can be proved by expansion.

0.4.1.1 Partial differentiation

This may be required when the vector is a function of more than one scalar independent variable. For instance, consider a vector $\mathbf{V}$ which is a function of the Cartesian coordinates x, y, z of a point in space. Then $(\partial\mathbf{V}/\partial x)$ is the rate of change of $\mathbf{V}$ with respect to x when y and z remain constant. Similarly for $(\partial\mathbf{V}/\partial y)$ and $(\partial\mathbf{V}/\partial z)$.

If x, y, and z change simultaneously by the differential increments δx, δy, and δz, the total differential $d\mathbf{V}$ of $\mathbf{V}$ will be

$$d\mathbf{V} = \left(\frac{\partial \mathbf{V}}{\partial x}\right)dx + \left(\frac{\partial \mathbf{V}}{\partial y}\right)dy + \left(\frac{\partial \mathbf{V}}{\partial z}\right)dz$$

Similarly if $\mathbf{r}$ is a radius vector from the origin, that is, $\mathbf{r} = \mathbf{i}_x x + \mathbf{i}_y y + \mathbf{i}_z z$, then its differential increment is given by $d\mathbf{r} = \mathbf{i}_x dx + \mathbf{i}_y dy + \mathbf{i}_z dz$. Thus $d\mathbf{V}$ can also be written as (symbolically):

$$d\mathbf{V} = \left[\left(\frac{\partial}{\partial x}\right)dx + \left(\frac{\partial}{\partial y}\right)dy + \left(\frac{\partial}{\partial z}\right)dz\right]\mathbf{V} \tag{0.32}$$

If the operator ∇ (called 'del' or 'nabla') is defined by

$$\nabla \equiv \mathbf{i}_x\left(\frac{\partial}{\partial x}\right) + \mathbf{i}_y\left(\frac{\partial}{\partial y}\right) + \mathbf{i}_z\left(\frac{\partial}{\partial z}\right) \qquad \text{(in Cartesian coordinates)} \tag{0.33}$$

then regarding it as a kind of vector, we get

$$d\mathbf{V} = (\nabla \cdot d\mathbf{r})\,\mathbf{V} \tag{0.34}$$

0.4.2 Integration—Line, Surface, and Volume Integrals

These are all extensions of the definition of the usual one-dimensional integral as the limit of a sum. Recapitulating, if a function $f(x)$ is defined over a range a to b of values of x, then the integral is given by

$$\int_a^b f(x)dx = \lim_{n\to\infty} \sum_{i=1}^{n} f_i\, \delta x_i \tag{0.35}$$

where the interval a to b has been subdivided into n continuous intervals of lengths δx_1, δx_2, …, δx_i, …, δx_n (Figure 0.9).

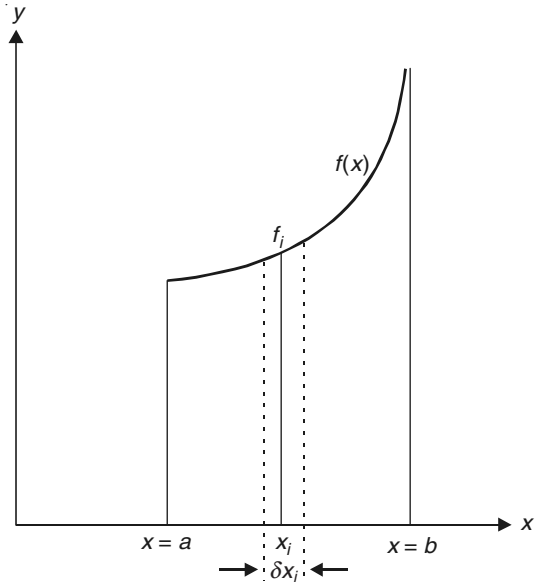


Figure 0.9 Integration of $f(x)$.

Similarly, we can also define the integral of a scalar function over a prescribed path in the three-dimensional space.

Line integral. Let C be a curve joining the points P and Q and $f(x, y, z)$ be a function of position defined over a region of space including C. Then the line integral of $f(x, y, z)$ over the curve C is

$$\int_{\mathrm{P}(s)}^{\mathrm{Q}(s)} f(x, y, z)\, ds = \lim_{n\to\infty} \sum_{i=1}^{n} f_i\, \delta s_i \tag{0.36}$$

where δs_i is the ith interval over the length PQ of the curve C, and f_i is the value of $f(x, y, z)$ at that interval. The requirement of continuity and the single-valued nature of $f(x, y, z)$ are also assumed. Thus for a vector function, say $\mathbf{B}$, over a path OA (Figure 0.10), the line integral will be

$$\int_{\mathrm{O}}^{\mathrm{A}} \mathbf{B}\cdot d\mathbf{s} = \int_{\mathrm{O}}^{\mathrm{A}} B\cos\theta\, ds$$

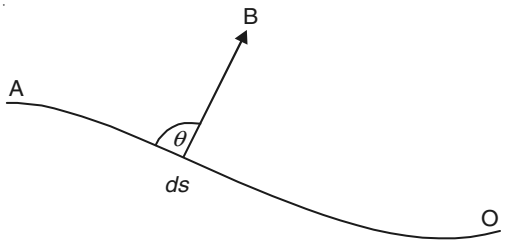


Figure 0.10 Line integral for a vector function.

When the path is a closed one, this integral is written as

$$\oint_C \mathbf{B} \cdot d\mathbf{s}$$

Surface integral. This is also defined along similar lines. Given a surface S and a function $f(x, y, z)$ defined in a region including S, the surface integral of $f(x, y, z)$ is an additive function of the elements of the area of S. As usual, S is subdivided into n contiguous elements (Figure 0.11) of areas δS_1, δS_2, …, δS_i, …, δS_n, and assuming that $f(x, y, z)$ is continuous and single-valued over a region including S, then the surface integral is given by

$$\iint_S f(x, y, z)\, dS = \lim_{n \to \infty} \sum_{i=1}^{n} f_i \delta S_i \tag{0.36a}$$

where f_i is the value of $f(x, y, z)$ at some point within the element of area δS_i.

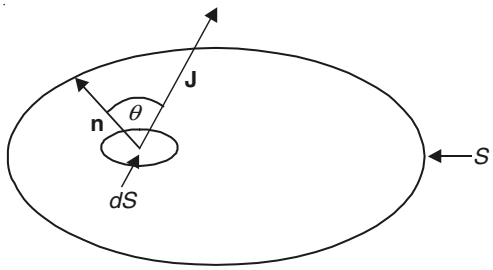


Figure 0.11 Surface integral.

In the most common types of surface integrals, the function $f(x, y, z)$ represents the normal component of the vector. For example, the total current flowing through a surface S can be expressed as a surface integral of the normal component J_n of the current density $\mathbf{J}$. That is,

$$\begin{aligned} I &= \iint_S J_n dS \\ &= \iint_S J\mathbf{n} \cdot d\mathbf{S} = \iint_S \mathbf{J} \cdot d\mathbf{S} \\ &= \iint_S J \cos\theta \, dS \end{aligned}$$

where $\mathbf{n}$ is the unit normal vector.

Usually for any vector the above integral is the flux of **J** through *S*. If *S* is a closed surface integral over *S*, it gives the total flux of **J** out of *S*, and is denoted by

$$\oiint_S \mathbf{J} \cdot d\mathbf{S}$$

In this case, the unit vector **n** is usually taken to be outwardly directed.

Volume integral. Similarly, the volume integral of a function $f(x, y, z)$ is an additive function of the volume of integration and can be obtained by considering similar differential elements δv_i of the volume v. This is denoted as

$$\iiint_v f(x, y, z)\, dv$$

A sign of circle is added to the sign of integration when the volume of integration under consideration is the entire space. As will be seen later, these three types of integrals can be interrelated.

0.5 THE VECTOR OPERATOR ∇ AND ITS USES

The differential operator ∇ was introduced by Sir William Rowan Hamilton and developed by P.G. Tait. It is of paramount importance in all three-dimensional physical problems. This symbol was originally named 'nabla', though it is now more commonly known as 'del' as introduced by J. Willard Gibbs. In Cartesian notation,

$$\text{del} = \nabla \equiv \mathbf{i}_x \left(\frac{\partial}{\partial x}\right) + \mathbf{i}_y \left(\frac{\partial}{\partial y}\right) + \mathbf{i}_z \left(\frac{\partial}{\partial z}\right)$$

and can be applied as a directive differentiator to either a scalar or a vector function of space. There are three fundamental operations with ∇ which are of physical interest.

0.5.1 The Gradient of a Scalar

This first operator pertains to scalar functions of space, which are of significant importance in the study of electromagnetic phenomena, for instance, electric potential.

Consider a scalar function *U*—a single-valued function of position. The spatial behaviour of *U* can be defined geometrically by means of a family of constant *U* surfaces. For example, if *U* is the distance from a given reference point, then the constant *U* surfaces are concentric spheres. The normal trajectories to this family of surfaces are the lines in the direction in which *U* changes most rapidly. Thus it should be possible to represent at each point, both the direction of the trajectory and the corresponding maximum rate of change of *U* by means of an appropriately defined vector. This vector is called the 'gradient of *U*' or 'grad *U*'.

From Figure 0.12,

$$\frac{dU}{dr} = \frac{dU}{dn} \cos\theta \qquad \text{and} \qquad dn = dr \cos\theta$$

Thus,

$$\text{grad}\, U = \mathbf{n} \frac{\partial U}{\partial n} \tag{0.37}$$

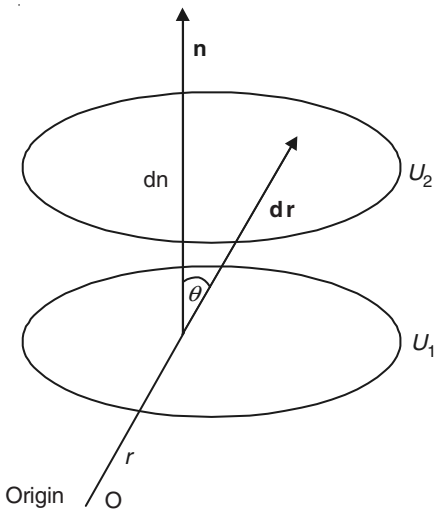


Figure 0.12 Gradient of a scalar.

Now in Cartesian coordinates, consider the vector ∇U, i.e.

$$\nabla U \equiv \mathbf{i}_x \left(\frac{\partial U}{\partial x}\right) + \mathbf{i}_y \left(\frac{\partial U}{\partial y}\right) + \mathbf{i}_z \left(\frac{\partial U}{\partial z}\right) \tag{0.38}$$

The vector rates of change of U in the directions of the coordinate axes are $\mathbf{i}_x(\partial U/\partial x)$, $\mathbf{i}_y(\partial U/\partial y)$, and $\mathbf{i}_z(\partial U/\partial z)$, respectively. Their sum will be a vector with magnitude and direction of the most rapid rate of change of U. It can be shown that this is equivalent to the vector 'gradient' defined as follows:

From the differential increment of the radius vector

$$\mathbf{dr} = \mathbf{i}_x dx + \mathbf{i}_y dy + \mathbf{i}_z dz$$

$$\therefore \quad (\text{grad } U)\cdot \mathbf{dr} = \left(\frac{\partial U}{\partial n}\right)\mathbf{n}\cdot\mathbf{dr} = \left(\frac{\partial U}{\partial n}\right) dr \cos\theta$$

$$= \left(\frac{\partial U}{\partial n}\right) dn = dU$$

and

$$dU = \left(\frac{\partial U}{\partial x}\right)dx + \left(\frac{\partial U}{\partial y}\right)dy + \left(\frac{\partial U}{\partial z}\right)dz$$

$$= \left\{\mathbf{i}_x \left(\frac{\partial U}{\partial x}\right) + \mathbf{i}_y \left(\frac{\partial U}{\partial y}\right) + \mathbf{i}_z \left(\frac{\partial U}{\partial z}\right)\right\} \cdot \left\{\mathbf{i}_x dx + \mathbf{i}_y dy + \mathbf{i}_z dz\right\}$$

$$= \nabla U \cdot \mathbf{dr} \tag{0.39}$$

$$\therefore \quad \text{grad } U = \nabla U \equiv \mathbf{i}_x \left(\frac{\partial U}{\partial x}\right) + \mathbf{i}_y \left(\frac{\partial U}{\partial y}\right) + \mathbf{i}_z \left(\frac{\partial U}{\partial z}\right)$$

Next, let $\mathbf{V}$ = grad $U = \nabla U$.

Taking the line integral of $\mathbf{V}$ from A to B (as shown in Figure 0.13),

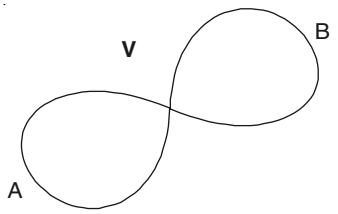


Figure 0.13 Line integral between two points.

$$\int_{A}^{B} \mathbf{V} \cdot d\mathbf{l} = \int_{A}^{B} (\text{grad}\, U) \cdot dr = \int_{A}^{B} dU = U_B - U_A$$

Thus the value of this integral is independent of the path of integration. Hence considering this integral over a closed path,

$$\oint_{ABA} (\text{grad}\, U) \cdot dl = \int + \int = (U_B - U_A) + (U_A - U_B) = 0 \tag{0.40}$$

Thus, when a vector field can be expressed as the gradient of a scalar field, the line integral of the vector between two points is independent of the path followed and equals the difference of the values of the scalar at its ends. Also, for such a vector field the line integral along any closed path is zero. A vector field so derived from the gradient relationship of a scalar is sometimes called a *scalar potential field*. This vector is also known as 'lamellar'.

0.5.1.1 Directional derivative

The quantity (**A.** grad S), when **A** is a unit vector, is called the *directional derivative* of the scalar S in the direction of the unit vector **A** and equals the rate of change of S in the direction of **A.** More generally when **A** is not necessarily a unit vector,

$$(\mathbf{A} \cdot \nabla) S = \mathbf{A} \cdot \text{grad}\, S = A_x \left(\frac{\partial S}{dx}\right) + A_y \left(\frac{\partial S}{dy}\right) + A_z \left(\frac{\partial S}{dz}\right)$$

The scalar operator ($\mathbf{A} \cdot$ grad) can also be applied to a vector **B**, giving

$$\begin{aligned} (\mathbf{A} \cdot \nabla)\, \mathbf{B} &= (\mathbf{A} \cdot \text{grad})\, \mathbf{B} \\ &= \mathbf{i}_x (\mathbf{A} \cdot \text{grad}\, B_x) + \mathbf{i}_y (\mathbf{A} \cdot \text{grad}\, B_y) + \mathbf{i}_z (\mathbf{A} \cdot \text{grad}\, B_z) \end{aligned} \tag{0.41}$$

giving in all nine terms.

0.5.2 The Divergence of a Vector

In Section 0.5.1, the del (∇) operator was operated on a scalar field to produce a vector function. Now we shall operate it (∇) on a vector field to produce a scalar function. So let us consider the field of a vector **A**. Over a closed surface S, enclosing a volume v, the outward flux of **A** (denoted as Φ) through S is

$$\Phi = \iint_S A_n \, dS = \iint_S \mathbf{A} \cdot \mathbf{n} \, dS \tag{0.42}$$

where **n** is the unit vector normal to *S*.

The divergence of the vector field **A** is defined as

$$\text{div } \mathbf{A} = \lim_{v \to 0} \left(\iint_S \frac{\mathbf{A} \cdot \mathbf{n} \, dS}{v} \right)$$

$$= \lim_{v \to 0} \frac{\text{flux of } \mathbf{A} \text{ out of } S}{\text{volume enclosed by } S} \tag{0.43}$$

Differential form of derivation of divergence in Cartesian coordinates. In the vector field mentioned above, consider an element of volume enclosed by the lengths Δx, Δy, Δz (parallel to the coordinate axes). (See Figure 0.14.)

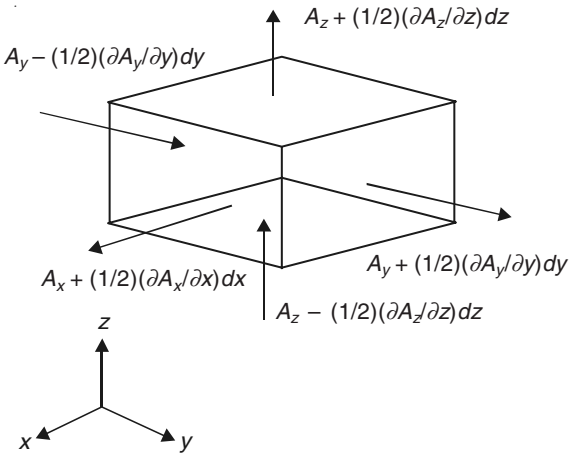


Figure 0.14 Divergence of a vector point function.

Let the value of the vector at the centre of this element box be

$$\mathbf{A} = \mathbf{i}_x A + \mathbf{i}_y A_y + \mathbf{i}_z A_z$$

Considering the two parallel faces ($\Delta y \Delta z$), the total flux of **A** through these two faces is

$$\left\{ A_x + \left(\frac{1}{2}\right)\left(\frac{\partial A_x}{\partial x}\right) dx \right\} dy\, dz - \left\{ A_x - \left(\frac{1}{2}\right)\left(\frac{\partial A_x}{\partial x}\right) dx \right\} dy\, dz = \frac{\partial A_x}{\partial x} dx\, dy\, dz$$

Similarly, the contribution to the flux of **A** from the remaining two pairs of parallel faces will be:

From the faces ($dz dx$): $\left(\dfrac{\partial A_y}{\partial y}\right) dx\, dy\, dz$

From the faces ($dx dy$): $\left(\dfrac{\partial A_z}{\partial z}\right) dx\, dy\, dz$

Thus, the total flux of **A** from the box element is

$$\left\{\left(\frac{\partial A_x}{\partial x}\right)+\left(\frac{\partial A_y}{\partial y}\right)+\left(\frac{\partial A_z}{\partial z}\right)\right\} dx\,dy\,dz$$

$$\therefore \quad \text{div } \mathbf{A} = \left\{\left(\frac{\partial A_x}{\partial x}\right)+\left(\frac{\partial A_y}{\partial y}\right)+\left(\frac{\partial A_z}{\partial z}\right)\right\}$$

$$= \left\{\mathbf{i}_x\left(\frac{\partial}{\partial x}\right)+\mathbf{i}_y\left(\frac{\partial}{\partial y}\right)+\mathbf{i}_z\left(\frac{\partial}{\partial z}\right)\right\}\cdot\left\{\mathbf{i}_x A_x+\mathbf{i}_y A_y+\mathbf{i}_z A_z\right\}$$

$$= \nabla\cdot\mathbf{A} \tag{0.44}$$

If no flux is coming out of a volume, then $\nabla\cdot\mathbf{A} = 0$.

Such a field is called 'Solenoidal' (from the Greek word 'Solen'= tube). Since div **A** is the flux of the vector **A**,

$$\oiint_S \mathbf{A}\cdot\mathbf{n}\,dS = \iiint_v \nabla\cdot\mathbf{A}\,dv \tag{0.45}$$

This is the 'Gauss' theorem', which stated in words is:

The flux of any vector quantity over any arbitrary closed surface is equal to the strength of the enclosed sources, that is, the flux of **A** *through S is equal to the volume integral of div* **A** *over the enclosed volume v.*

The Gauss' theorem is the basis for the integral definition of divergence.

0.5.3 The Curl of a Vector

This is the third application of the del operator where it is operated on a vector field such that another vector field is obtained. In a vector field **A** there is a point P about which we describe a closed path *C* enclosing an area *S* (Figure 0.15).

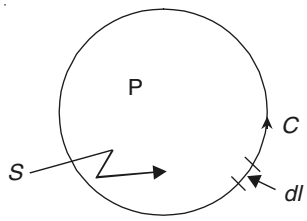


Figure 0.15 Contour for the curl.

Consider now the line integral of **A** over this contour. The direction of integration is such that the area *S* enclosed by the contour is kept to the left during the traverse of the path. A new vector 'curl of **A**' (≡ curl **A**) is defined such that its magnitude is

$$|\text{curl } \mathbf{A}| = \lim_{S\to 0}\oint_C \mathbf{A}\cdot\mathbf{dl} \tag{0.46}$$

Since curl **A** is a vector, its direction is perpendicular to the plane of C with the positive sense fixed with respect to the direction of integration by the right-hand screw rule. Hence it can have components in the x, y, and z directions.

Derivation of curl in rectangular Cartesian coordinates. Any area as discussed above can be resolved into components in the three planes defined by the coordinate axes . Also, it should be noted that when the area under consideration is perpendicular to the vector field, none of the field is directed along any part of the bounding edge of the area and hence the line integral round it is zero. When the area under consideration is parallel to the field, the value of the vector around the boundary may change and the line integral can have a finite nonzero value. Consider the area PQRS of sides δx and δy (Figure 0.16).

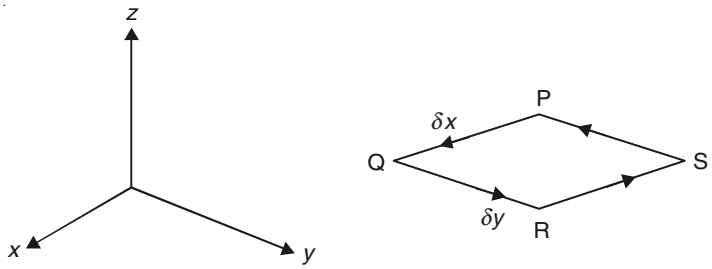


Figure 0.16 Contour for the line integral for deriving the curl.

Let the value of the vector **A** at the centre of the rectangle PQRS be:

$$\mathbf{A} = \mathbf{i}_x A_x + \mathbf{i}_y A_y + \mathbf{i}_z A_z$$

Contributions to the line integral along PQ and RS from the x-component of **A** have the values:

$$A_x - \left(\frac{1}{2}\right)\left(\frac{\partial A_x}{\partial y}\right)\delta y \quad \text{along PQ}$$

and

$$A_x + \left(\frac{1}{2}\right)\left(\frac{\partial A_x}{\partial y}\right)\delta y \quad \text{along RS}$$

and the contributions along QR and SP have the values:

$$A_y + \left(\frac{1}{2}\right)\left(\frac{\partial A_y}{\partial x}\right)\delta x \quad \text{along QR}$$

$$A_y - \left(\frac{1}{2}\right)\left(\frac{\partial A_y}{\partial x}\right)\delta x \quad \text{along SP}$$

$\therefore$ The line integral along PQRS will be

$$\left[\left\{A_x - \left(\frac{1}{2}\right)\left(\frac{\partial A_x}{\partial y}\right)\delta y\right\} - \left\{A_x + \left(\frac{1}{2}\right)\left(\frac{\partial A_x}{\partial y}\right)\delta y\right\}\right]\delta x \; +$$

$$\left[\left\{A_y + \left(\frac{1}{2}\right)\left(\frac{\partial A_y}{\partial x}\right)\delta x\right\} - \left\{A_y - \left(\frac{1}{2}\right)\left(\frac{\partial A_y}{\partial x}\right)\delta x\right\}\right]\delta y$$

$$= \left\{\left(\frac{\partial A_y}{\partial x}\right) - \left(\frac{\partial A_x}{\partial y}\right)\right\}\delta x \delta y$$

The area of the element is $\delta x \delta y$.

$$\therefore \qquad z\text{-component of curl } \mathbf{A} = \left\{\left(\frac{\partial A_y}{\partial x}\right) - \left(\frac{\partial A_x}{\partial y}\right)\right\}$$

Similarly,

$$y\text{-component of curl } \mathbf{A} = \left\{\left(\frac{\partial A_x}{\partial z}\right) - \left(\frac{\partial A_z}{\partial x}\right)\right\}$$

and

$$x\text{-component of curl } \mathbf{A} = \left\{\left(\frac{\partial A_z}{\partial y}\right) - \left(\frac{\partial A_y}{\partial z}\right)\right\}$$

$$\therefore \qquad \text{curl } \mathbf{A} = \mathbf{i}_x\left\{\left(\frac{\partial A_z}{\partial y}\right) - \left(\frac{\partial A_y}{\partial z}\right)\right\} + \mathbf{i}_y\left\{\left(\frac{\partial A_x}{\partial z}\right) - \left(\frac{\partial A_z}{\partial x}\right)\right\}$$

$$+ \mathbf{i}_z\left\{\left(\frac{\partial A_y}{\partial x}\right) - \left(\frac{\partial A_x}{\partial y}\right)\right\}$$

$$= \begin{bmatrix} \mathbf{i}_x & \mathbf{i}_y & \mathbf{i}_z \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ A_x & A_y & A_z \end{bmatrix}$$

Using the del operator,

$$\nabla \times \mathbf{A} = \left\{\mathbf{i}_x\left(\frac{\partial}{\partial x}\right) + \mathbf{i}_y\left(\frac{\partial}{\partial y}\right) + \mathbf{i}_z\left(\frac{\partial}{\partial z}\right)\right\} \times \left\{\mathbf{i}_x A_x + \mathbf{i}_y A_y + \mathbf{i}_z A_z\right\}$$

$$= \text{curl } \mathbf{A} \tag{0.47}$$

The physical meaning of the curl vector may be seen as the measure of the rotational capability or the vorticity of the original vector. A very important property of the field of curl **A** is that it is purely rotational, i.e. its divergence is identically zero. Thus,

$$\text{div curl } \mathbf{A} = 0 \tag{0.48}$$

This property will be checked in a later section by performing the required operations on the Cartesian components of **A**. Another important property of the curl operation follows from its integral definition and the line integral of the grad vector which over any closed path is

$$\text{curl grad } U = 0 \tag{0.49}$$

This is a vector identity. Such a vector field whose curl is zero is called a 'conservative' field. It is so called because the energy is conserved in a force field having this property, for any motion that ends at the point of origin, for instance, the net work done by an electrostatic field on a unit charge is equal to zero when the charge is returned to its point of origin. Hence the work done by the field in moving the charge from one point to another is independent of the path followed. In mathematical terms, a conservative vector can be represented as the negative gradient of a scalar potential function.

0.6 SOME INTEGRAL THEOREMS OF VECTORS

Initially recapitulating:

0.6.1 Gauss' Theorem

For the vector **A**

$$\oiint_S (\mathbf{A}\cdot\mathbf{n})\,dS = \oiint_S \mathbf{A}\cdot d\mathbf{S} = \iiint_v (\text{div }\mathbf{A})\,dv \left\{ \equiv \iiint_v (\nabla\cdot\mathbf{A})\,dv \right\} \tag{0.50}$$

where v is the volume enclosed by the closed surface S. Gauss' theorem relates the surface integral of **A** with the volume integral of $\nabla\cdot\mathbf{A}$.

0.6.2 Stokes' Theorem

For a vector **A**

$$\iint_S (\text{curl }\mathbf{A})\cdot dS = \iint_S (\nabla\times\mathbf{A})\cdot d\mathbf{S} = \oint_C \mathbf{A}\,d\mathbf{l} \tag{0.51}$$

where S is the surface enclosed by the closed curve C. Stokes' theorem relates the line integral of **A** with the surface integral of $\nabla\times\mathbf{A}$.

0.6.3 Green's Theorem

Let U and V be two scalar functions which together with their gradients and Laplacians are non-singular within a volume v and on the enclosing surface S, that is, continuous and finite in the region of integration, and can be differentiated twice. Hence applying the Gauss' theorem to the vector (V grad $U = V(\nabla U)$), we have

$$\iiint_v \text{div}\{V\ \text{grad}\ U\}dv \;=\; \oiint_S \{V\ \text{grad}\ U\}\cdot d\mathbf{S} \tag{0.52}$$

But

$$\text{div}\ \{V\ \text{grad}\ U\} = \text{grad}\ V\cdot\text{grad}\ U + V\nabla^2 U \qquad \text{(Refer to Section 0.7)} \tag{0.53}$$

Hence substituting Eq. (0.53) in Eq. (0.52), we obtain

$$\iiint_v \left\{(\text{grad}\ V)\cdot(\text{grad}\ U) + V\nabla^2 U\right\} dv \;=\; \oiint_S \{V\ \text{grad}\ U\}\cdot d\mathbf{S}$$

$$= \oiint_S \left\{V\left(\partial U/\partial n\right)\right\} dS \tag{0.54}$$

This equation is known as the 'Green's first identity'. By considering similar operations on the other vector $\{U \text{ grad } V\}$ obtained by interchanging U and V, we obtain

$$\iiint_v \left\{(\text{grad } U)\cdot(\text{grad } V) + U\nabla^2 V\right\} dv = \oiint_S \{U \text{ grad } V\}\cdot d\mathbf{S}$$

$$= \oiint_S \left\{U\left(\partial V/\partial n\right)\right\} dS \tag{0.55}$$

Subtracting Eq. (0.55) from Eq. (0.54), we have

$$\iiint_v \left\{V\left(\nabla^2 U\right) - U\left(\nabla^2 V\right)\right\} dv = \oiint_S \left\{V\left(\text{grad } U\right) - U\left(\text{grad } V\right)\right\}\cdot d\mathbf{S}$$

$$= \oiint_S \left\{V\left(\partial U/\partial n\right) - U\left(\partial V/\partial n\right)\right\} dS \tag{0.56}$$

This is known as the 'Green's second identity' or the 'Green's theorem'. The above surface integral is of the outward normal component of the vector within the brackets over the closed surface S and the volume integral is of the scalar quantity within the brackets on the left-hand side over the volume v enclosed by the surface S. This relationship between the surface integral and the volume integral is called the 'Green's theorem'. It is of use in deriving the properties of the 'Green's functions' and provides a mathematical basis for the 'method of images'.

0.6.4 Vector Analogue of Green's Theorem

A useful vector analogue of Eqs. (0.54) and (0.55) has been proved by Stratton. In Section 0.6.3, the starting point was the vector (V (grad U)), which was built from the two scalar functions U and V. Now we have two vector functions $\mathbf{P}$ and $\mathbf{Q}$ both of which are finite and continuous in the region of integration and can be differentiated twice, and started with the vector field ($\mathbf{P} \times (\nabla \times \mathbf{Q})$), that is, a vector which is the cross product of the vector $\mathbf{P}$ with the curl of the second vector $\mathbf{Q}$. Let us now consider the divergence of this new vector to which we again apply the Gauss' theorem. Thus,

$$\iiint_v \left[\nabla\cdot\left\{\mathbf{P}\times\left(\nabla\times\mathbf{Q}\right)\right\}\right] dv = \oiint_S \left\{\mathbf{P}\times\left(\nabla\times\mathbf{Q}\right)\right\}\cdot d\mathbf{S} \tag{0.57}$$

But

$$\nabla \cdot \{\mathbf{P} \times (\nabla \times \mathbf{Q})\} = (\nabla \times \mathbf{Q}) \cdot (\nabla \times \mathbf{P}) - \mathbf{P} \cdot (\nabla \times \nabla \times \mathbf{Q})\} \quad \text{(Refer to Section 0.7)} \tag{0.58}$$

Hence, we obtain

$$\iiint_v \left\{(\nabla\times\mathbf{Q})\cdot(\nabla\times\mathbf{P}) - \mathbf{P}\cdot(\nabla\times\nabla\times\mathbf{Q})\right\} dv = \oiint_S \left\{\mathbf{P}\times\left(\nabla\times\mathbf{Q}\right)\right\}\cdot d\mathbf{S} \tag{0.59}$$

By interchanging the vectors $\mathbf{P}$ and $\mathbf{Q}$ as before,

$$\iiint_v \left\{(\nabla\times\mathbf{P})\cdot(\nabla\times\mathbf{Q}) - \mathbf{Q}\cdot(\nabla\times\nabla\times\mathbf{P})\right\} dv = \oiint_S \left\{\mathbf{Q}\times\left(\nabla\times\mathbf{P}\right)\right\}\cdot d\mathbf{S} \tag{0.60}$$

Subtracting Eq. (0.60) from Eq. (0.59), we obtain

$$\iiint_v \{\mathbf{Q}\cdot(\nabla\times\nabla\times\mathbf{P}) - \mathbf{P}\cdot(\nabla\times\nabla\times\mathbf{Q})\}dv$$

$$= \oiint_S \left[\{\mathbf{P}\times(\nabla\times\mathbf{Q})\} - \{\mathbf{Q}\times(\nabla\times\mathbf{P})\}\right]\cdot d\mathbf{S} \tag{0.61}$$

This is the vector analogue of the Green's theorem, that is, Eq. (0.56). Its usefulness lies in the derivation of the images of magnetic systems. The images in magnetic systems are somewhat more complicated than those of electrostatic systems because the sources in electrostatic systems are charges which can be represented as scalar quantities, whereas the sources of the magnetic fields are in terms of currents which have to be represented vectorially. We will discuss these points in depth while dealing with them later.

0.7 APPLICATIONS OF THE OPERATOR 'DEL' ($= \nabla$)

0.7.1 The Operator Div Grad

If S is a scalar function of position in space, then grad $S = \nabla S$ is a vector function which gives both the magnitude and the direction of the greatest rate of change of S. Since this is a vector, it can have a divergence, that is,

$$\text{div grad } S = \nabla\cdot(\nabla S) = \nabla\cdot\nabla S = \nabla^2 S \tag{0.62}$$

In Cartesian coordinates,

$$\text{div grad } S = \left\{\mathbf{i}_x\left(\frac{\partial}{\partial x}\right) + \mathbf{i}_y\left(\frac{\partial}{\partial y}\right) + \mathbf{i}_z\left(\frac{\partial}{\partial z}\right)\right\}\cdot\left\{\mathbf{i}_x\left(\frac{\partial S}{\partial x}\right) + \mathbf{i}_y\left(\frac{\partial S}{\partial y}\right) + \mathbf{i}_z\left(\frac{\partial S}{\partial z}\right)\right\}$$

$$= \frac{\partial^2 S}{\partial x^2} + \frac{\partial^2 S}{\partial y^2} + \frac{\partial^2 S}{\partial z^2} \tag{0.63}$$

and the operator

$$\text{div grad} \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \equiv \nabla^2 \tag{0.64}$$

is called the Laplace's operator and has many applications in electromagnetism.

0.7.2 Divergence of a Vector Product

This operation is required in calculating the flux of energy in an electromagnetic field. Let **A** and **B** be two vectors. Hence in the Cartesian coordinate system, the vector product **V** is:

$$\mathbf{V} = \mathbf{A}\times\mathbf{B} = \mathbf{i}_x(A_yB_z - A_zB_y) + \mathbf{i}_y(A_zB_x - A_xB_z) + \mathbf{i}_z(A_xB_y - A_yB_x)$$

$$= \mathbf{i}_xV_x + \mathbf{i}_yV_y + \mathbf{i}_zV_z \tag{0.65}$$

Hence,

$$\text{div } \mathbf{V} = \nabla \cdot \mathbf{V} = \left\{\left(\frac{\partial V_x}{\partial x}\right)+\left(\frac{\partial V_y}{\partial y}\right)+\left(\frac{\partial V_z}{\partial z}\right)\right\}$$

$$= \left(\frac{\partial A_y}{\partial x}\right)B_z + A_y\left(\frac{\partial B_z}{\partial x}\right)-\left(\frac{\partial A_z}{\partial x}\right)B_y - A_z\left(\frac{\partial B_y}{\partial x}\right)$$

$$+\left(\frac{\partial A_z}{\partial y}\right)B_x + A_z\left(\frac{\partial B_x}{\partial y}\right)-\left(\frac{\partial A_x}{\partial y}\right)B_z - A_x\left(\frac{\partial B_z}{\partial y}\right)$$

$$+\left(\frac{\partial A_x}{\partial z}\right)B_y + A_x\left(\frac{\partial B_y}{\partial z}\right)-\left(\frac{\partial A_y}{\partial z}\right)B_x - A_y\left(\frac{\partial B_x}{\partial z}\right)$$

$$= \left\{B_x\left[\left(\frac{\partial A_z}{\partial y}\right)-\left(\frac{\partial A_y}{\partial z}\right)\right] + B_y\left[\left(\frac{\partial A_x}{\partial z}\right)-\left(\frac{\partial A_z}{\partial x}\right)\right]\right.$$

$$\left.+ B_z\left[\left(\frac{\partial A_y}{\partial x}\right)-\left(\frac{\partial A_x}{\partial y}\right)\right]\right\}$$

$$-\left\{A_x\left[\left(\frac{\partial B_z}{\partial y}\right)-\left(\frac{\partial B_y}{\partial z}\right)\right] + A_y\left[\left(\frac{\partial B_x}{\partial z}\right)-\left(\frac{\partial B_z}{\partial x}\right)\right]\right.$$

$$\left.+ A_z\left[\left(\frac{\partial B_y}{\partial x}\right)-\left(\frac{\partial B_x}{\partial y}\right)\right]\right\}$$

$$= \mathbf{B} \cdot \text{curl } \mathbf{A} - \mathbf{A} \cdot \text{curl } \mathbf{B} = \mathbf{B} \cdot (\nabla \times \mathbf{A}) - \mathbf{A} \cdot (\nabla \times \mathbf{B})$$

or

$$\nabla \cdot (\mathbf{A} \times \mathbf{B}) = \mathbf{B} \cdot (\nabla \times \mathbf{A}) - \mathbf{A} \cdot (\nabla \times \mathbf{B}) \tag{0.66}$$

0.7.3 Divergence and Curl of $S\mathbf{A}$

Consider a vector field which itself is the product of a scalar field S and a vector field $\mathbf{A}$ at every point. Hence,

$$\text{div } (S\mathbf{A}) = \nabla \cdot (S\mathbf{A})$$

$$= \frac{\partial(SA_x)}{\partial x} + \frac{\partial(SA_y)}{\partial y} + \frac{\partial(SA_z)}{\partial z}$$

$$= S\left(\frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z}\right)+ A_x\left(\frac{\partial S}{\partial x}\right)+ A_y\left(\frac{\partial S}{\partial y}\right)+ A_z\left(\frac{\partial S}{\partial z}\right)$$

$$= S \text{ div } \mathbf{A} + \left(\mathbf{i}_x A_x + \mathbf{i}_y A_y + \mathbf{i}_z A_z\right) \cdot \left\{\mathbf{i}_x \left(\frac{\partial S}{\partial x}\right) + \mathbf{i}_y \left(\frac{\partial S}{\partial y}\right) + \mathbf{i}_z \left(\frac{\partial S}{\partial z}\right)\right\}$$

$$= S \text{ div } \mathbf{A} + \mathbf{A} \cdot \text{grad } S$$

$$= S\,(\nabla \cdot \mathbf{A}) + \mathbf{A} \cdot (\nabla S)$$

$\therefore$
$$\nabla \cdot (S\mathbf{A}) = S(\nabla \cdot \mathbf{A}) + \mathbf{A} \cdot (\nabla S) \tag{0.67}$$

Next

$$\text{curl } (S \cdot \mathbf{A}) = \nabla \times (S\mathbf{A})$$

$$= \begin{bmatrix} \mathbf{i}_x & \mathbf{i}_y & \mathbf{i}_z \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ SA_x & SA_y & SA_z \end{bmatrix}$$

$\therefore$
$$\{\text{curl } (S\mathbf{A})\}_x = \frac{\partial}{\partial y}\left(SA_z\right) - \frac{\partial}{\partial z}\left(SA_y\right)$$

$$= S\left[\left(\frac{\partial A_z}{\partial y}\right) - \left(\frac{\partial A_y}{\partial z}\right)\right] + \left[A_z \left(\frac{\partial S}{\partial y}\right) - A_y \left(\frac{\partial S}{\partial z}\right)\right]$$

Similarly,

$$\{\text{curl } (S\mathbf{A})\}_y = S\left[\left(\frac{\partial A_x}{\partial z}\right) - \left(\frac{\partial A_z}{\partial x}\right)\right] + \left[A_x \left(\frac{\partial S}{\partial z}\right) - A_z \left(\frac{\partial S}{\partial x}\right)\right]$$

and

$$\{\text{curl } (S\mathbf{A})\}_z = S\left[\left(\frac{\partial A_y}{\partial x}\right) - \left(\frac{\partial A_x}{\partial y}\right)\right] + \left[A_y \left(\frac{\partial S}{\partial x}\right) - A_x \left(\frac{\partial S}{\partial y}\right)\right]$$

$\therefore$ $\text{curl } (S\mathbf{A}) = S \text{ curl } \mathbf{A} + (\text{grad } S)\nabla \times \mathbf{A} = S \text{ curl } \mathbf{A} - \mathbf{A} \times \text{grad } S$

Hence,

$$\nabla \times (S\mathbf{A}) = S(\nabla \times \mathbf{A}) - \mathbf{A} \times (\nabla S) \tag{0.68}$$

0.7.4 The Operator Curl Grad

Grad S is a vector and it is possible to interpret the operator (curl grad) as being applied to a scalar point-function. In Cartesian coordinates,

$$\text{curl grad } S = \text{curl}\left\{\mathbf{i}_x \left(\frac{\partial S}{\partial x}\right) + \mathbf{i}_y \left(\frac{\partial S}{\partial y}\right) + \mathbf{i}_z \left(\frac{\partial S}{\partial z}\right)\right\}$$

$$= \begin{bmatrix} \mathbf{i}_x & \mathbf{i}_y & \mathbf{i}_z \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ \dfrac{\partial S}{\partial x} & \dfrac{\partial S}{\partial y} & \dfrac{\partial S}{\partial z} \end{bmatrix}$$

$$= \mathbf{i}_x \left[\frac{\partial}{\partial y}\left(\frac{\partial S}{\partial z}\right) - \frac{\partial}{\partial z}\left(\frac{\partial S}{\partial y}\right)\right]$$

$$+ \mathbf{i}_y \left[\frac{\partial}{\partial x}\left(\frac{\partial S}{\partial z}\right) - \frac{\partial}{\partial z}\left(\frac{\partial S}{\partial x}\right)\right]$$

$$+ \mathbf{i}_z \left[\frac{\partial}{\partial x}\left(\frac{\partial S}{\partial y}\right) - \frac{\partial}{\partial y}\left(\frac{\partial S}{\partial x}\right)\right]$$

$$= 0 \ \{= \nabla \times (\nabla S)\} \tag{0.69}$$

This is a vector identity and is the characteristic of a lamellar field, that is, the line integral of the scalar function S around any closed path is zero. Note that the scalar S is the scalar potential of the lamellar vector field (∇S).

0.7.5 The Operator ∇^2 with Vector Operand

A vector point function **V** can be expressed in terms of its rectangular components, and then each component can be considered as a scalar function of position, i.e.

$$\nabla^2 \mathbf{V} = \mathbf{i}_x \nabla^2 V_x + \mathbf{i}_y \nabla^2 V_y + \mathbf{i}_z \nabla^2 V_z \tag{0.70}$$

This result is of great importance in electromagnetics and hydromechanics.

0.7.6 The Operator Grad Div

For a vector field **V**, (div **V**) is a scalar field and hence has a gradient, which will be a vector. This new vector field, of which (div **V**) is the potential, is lamellar because curl grad is zero. Thus,

$$\text{grad div } \mathbf{V} = \nabla\,(\nabla \cdot \mathbf{V})$$

$$= \left\{ \mathbf{i}_x \left(\frac{\partial}{\partial x}\right) + \mathbf{i}_y \left(\frac{\partial}{\partial y}\right) + \mathbf{i}_z \left(\frac{\partial}{\partial z}\right)\right\} \left\{ \left(\frac{\partial V_x}{\partial x}\right) + \left(\frac{\partial V_y}{\partial y}\right) + \left(\frac{\partial V_z}{\partial z}\right)\right\}$$

$$= \mathbf{i}_x \left\{ \frac{\partial^2 V_x}{\partial x^2} + \frac{\partial^2 V_y}{\partial x\,\partial y} + \frac{\partial^2 V_z}{\partial x\,\partial z}\right\} + \mathbf{i}_y \left\{ \frac{\partial^2 V_x}{\partial x\,\partial y} + \frac{\partial^2 V_y}{\partial y^2} + \frac{\partial^2 V_z}{\partial y\,\partial z}\right\}$$

$$+ \mathbf{i}_z \left\{ \frac{\partial^2 V_x}{\partial x\,\partial z} + \frac{\partial^2 V_y}{\partial y\,\partial z} + \frac{\partial^2 V_z}{\partial z^2}\right\} \tag{0.71}$$

0.7.7 The Operator Div Curl

For any vector field **V**, curl **V** will also be a vector field whose divergence can be evaluated. Hence,

$$\text{div curl } \mathbf{V} = \nabla \cdot (\nabla \times \mathbf{V})$$

$$= \left[\left(\frac{\partial}{\partial x}\right)\left\{\left(\frac{\partial V_z}{\partial y}\right)-\left(\frac{\partial V_y}{\partial z}\right)\right\}+\left(\frac{\partial}{\partial y}\right)\left\{\left(\frac{\partial V_x}{\partial z}\right)-\left(\frac{\partial V_z}{\partial x}\right)\right\}\right.$$

$$\left.+\left(\frac{\partial}{\partial z}\right)\left\{\left(\frac{\partial V_y}{\partial x}\right)-\left(\frac{\partial V_x}{\partial y}\right)\right\}\right] = 0 \tag{0.72}$$

This again is a vector identity. Thus a vector field which is the curl of another vector field, is always a Solenoidal field. As its divergence is zero, it has no sources.

0.7.8 The Operator Curl Curl

From the expression of (curl **V**), the *x*-component of this operator will be:

$$(\text{curl curl } \mathbf{V})_x = \mathbf{i}_x\left[\left(\frac{\partial}{\partial y}\right)\left\{\left(\frac{\partial V_y}{\partial x}\right)-\left(\frac{\partial V_x}{\partial y}\right)\right\}-\left(\frac{\partial}{\partial z}\right)\left\{\left(\frac{\partial V_x}{\partial z}\right)-\left(\frac{\partial V_z}{\partial x}\right)\right\}\right]$$

$$= \mathbf{i}_x\left[\left\{\left(\frac{\partial^2 V_x}{\partial x^2}\right)+\left(\frac{\partial^2 V_y}{\partial x\,\partial y}\right)+\left(\frac{\partial^2 V_z}{\partial x\,\partial z}\right)\right\}-\left\{\left(\frac{\partial^2 V_x}{\partial x^2}\right)+\left(\frac{\partial^2 V_x}{\partial y^2}\right)+\left(\frac{\partial^2 V_x}{\partial z^2}\right)\right\}\right]$$

The above two expressions are the *x*-components of (grad div **V**) and $\nabla^2\mathbf{V}$. Hence considering all the three components:

$$\text{curl curl } \mathbf{V} = \nabla \times \nabla \times \mathbf{V} = \text{grad div } \mathbf{V} - \nabla^2\mathbf{V} \tag{0.73}$$

The operator equivalence is

$$\nabla \times \nabla \times \equiv \nabla\,(\nabla \cdot) - \nabla^2 \tag{0.74}$$

The above operator has applications in many electromagnetic problems.

0.7.9 Gradient of a Scalar Product

The scalar product *S* of the two vectors **A** and **B** in the Cartesian coordinate system is:

$$S = \mathbf{A}\cdot\mathbf{B} = A_xB_x + A_yB_y + A_zB_z \tag{0.11}$$

Hence,

$$\text{grad } S = \nabla S = \nabla(\mathbf{A}\cdot\mathbf{B})$$

$$= \mathbf{i}_x\left\{\left(\frac{\partial A_x}{\partial x}\right)B_x + A_x\left(\frac{\partial B_x}{\partial x}\right)+\left(\frac{\partial A_y}{\partial x}\right)B_y + A_y\left(\frac{\partial B_y}{\partial x}\right)+\left(\frac{\partial A_z}{\partial x}\right)B_z + A_z\left(\frac{\partial B_z}{\partial x}\right)\right\}$$

$$+ \mathbf{i}_y\left\{\left(\frac{\partial A_x}{\partial y}\right)B_x + A_x\left(\frac{\partial B_x}{\partial y}\right)+\left(\frac{\partial A_y}{\partial y}\right)B_y + A_y\left(\frac{\partial B_y}{\partial y}\right)+\left(\frac{\partial A_z}{\partial y}\right)B_z + A_z\left(\frac{\partial B_z}{\partial y}\right)\right\}+$$

$$\mathbf{i}_z\left\{\left(\frac{\partial A_x}{\partial z}\right)B_x + A_x\left(\frac{\partial B_x}{\partial z}\right)+\left(\frac{\partial A_y}{\partial z}\right)B_y + A_y\left(\frac{\partial B_y}{\partial z}\right)+\left(\frac{\partial A_z}{\partial z}\right)B_z + A_z\left(\frac{\partial B_z}{\partial z}\right)\right\}$$

From Section 0.5.1.1, we have the 'directional derivative' scalar operator{i.e. $(\mathbf{A}\cdot\nabla)$ or $(\mathbf{B}\cdot\nabla)$} operating on a vector, giving

$$(\mathbf{A}\cdot\nabla)\mathbf{B} = \mathbf{i}_x(\mathbf{A}\cdot\text{grad } B_x) + \mathbf{i}_y(\mathbf{A}\cdot\mathbf{grad}\ B_y) + \mathbf{i}_z(\mathbf{A}\cdot\text{grad } B_z)$$

$$= \mathbf{i}_x\left\{A_x\frac{\partial B_x}{\partial x} + A_y\frac{\partial B_x}{\partial y} + A_z\frac{\partial B_x}{\partial z}\right\}$$

$$+ \mathbf{i}_y\left\{A_x\frac{\partial B_y}{\partial x} + A_y\frac{\partial B_y}{\partial y} + A_z\frac{\partial B_y}{\partial z}\right\}$$

$$+ \mathbf{i}_z\left\{A_x\frac{\partial B_z}{\partial x} + A_y\frac{\partial B_z}{\partial y} + A_z\frac{\partial B_z}{\partial z}\right\}$$

and similarly

$$(B\cdot\nabla)\,\mathbf{A} = \mathbf{i}_x(\mathbf{B}\cdot\text{grad } A_x) + \mathbf{i}_y(\mathbf{B}\cdot\text{grad } A_y) + \mathbf{i}_z(\mathbf{B}\cdot\text{grad } A_z)$$

$$= \mathbf{i}_x\left\{B_x\frac{\partial A_x}{\partial x} + B_y\frac{\partial A_x}{\partial y} + B_z\frac{\partial A_x}{\partial z}\right\}$$

$$+ \mathbf{i}_y\left\{B_x\frac{\partial A_y}{\partial x} + B_y\frac{\partial A_y}{\partial y} + B_z\frac{\partial A_y}{\partial z}\right\}$$

$$+ \mathbf{i}_z\left\{B_x\frac{\partial A_z}{\partial x} + B_y\frac{\partial A_z}{\partial y} + B_z\frac{\partial A_z}{\partial z}\right\}$$

So, if we rewrite $\nabla(\mathbf{A}\cdot\mathbf{B})$ in terms of the above two quantities we will find that each component of this gradient will be:

$$\{\nabla(\mathbf{A}\cdot\mathbf{B})\}_x = \mathbf{i}_x\left[(\mathbf{A}\cdot\text{grad } B_x) - A_y\frac{\partial B_x}{\partial y} - A_z\frac{\partial B_x}{\partial z} + (\mathbf{B}.\text{grad } A_x) - B_y\frac{\partial A_x}{\partial y} - B_z\frac{\partial A_x}{\partial z}\right.$$

$$\left.+\left(\frac{\partial A_y}{\partial x}\right)B_y + A_y\left(\frac{\partial B_y}{\partial x}\right)+\left(\frac{\partial A_z}{\partial x}\right)B_z + A_z\left(\frac{\partial B_z}{\partial x}\right)\right]$$

$$= \mathbf{i}_x\left[\left(\mathbf{A}\cdot\text{grad } B_x\right) + A_y\left\{\frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y}\right\} + A_z\left\{\frac{\partial B_z}{\partial x} - \frac{\partial B_x}{\partial z}\right\}\right.$$

$$\left.+ (\mathbf{B}\cdot\text{grad } A_x) + B_y\left\{\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right\} + B_z\left\{\frac{\partial A_z}{\partial x} - \frac{\partial A_x}{\partial z}\right\}\right]$$

$$= \mathbf{i}_x(\mathbf{A}\cdot\text{grad } B_x) + (\mathbf{A}\times\text{curl }\mathbf{B})_x + \mathbf{i}_x(\mathbf{B}\cdot\text{grad } A_x) + (\mathbf{B}\times\text{curl }\mathbf{A})_x$$

Similarly for y- and z-components, i.e.

$$\{\nabla(\mathbf{A}\cdot\mathbf{B})\}_y = \mathbf{i}_y(\mathbf{A}\cdot\text{grad}\,B_y) + (\mathbf{A}\times\text{curl}\,\mathbf{B})_y + \mathbf{i}_y(\mathbf{B}\cdot\text{grad}\,A_y) + (\mathbf{B}\times\text{curl}\,\mathbf{A})_y$$

and

$$\{\nabla(\mathbf{A}\cdot\mathbf{B})\}_z = \mathbf{i}_z(\mathbf{A}\cdot\text{grad}\,B_z) + (\mathbf{A}\times\text{curl}\,\mathbf{B})_z + \mathbf{i}_z(\mathbf{B}\cdot\text{grad}\,A_z) + (\mathbf{B}\times\text{curl}\,\mathbf{A})_z$$

Hence, adding these three components,

$$\begin{aligned}\nabla(\mathbf{A}\cdot\mathbf{B}) &= (A\cdot\text{grad})\,\mathbf{B} + \mathbf{A}\times\text{curl}\,\mathbf{B} + (\mathbf{B}\cdot\text{grad})\,\mathbf{A} + \mathbf{B}\times\text{curl}\,\mathbf{A}\\ &= (\mathbf{A}\cdot\nabla)\mathbf{B} + \mathbf{A}\times(\nabla\times\mathbf{B}) + (\mathbf{B}\cdot\nabla)\,\mathbf{A} + \mathbf{B}\times(\nabla\times\mathbf{A})\end{aligned} \tag{0.74a}$$

0.7.10 Curl of a Vector Product

The vector product of the two vectors **A** and **B** in the Cartesian coordinate system is:

$$\begin{aligned}\mathbf{C} &= \mathbf{A}\times\mathbf{B}\\ &= \mathbf{i}_x(A_yB_z - A_zB_y) + \mathbf{i}_y(A_zB_x - A_xB_z) + \mathbf{i}_z(A_xB_y - A_yB_x)\end{aligned} \tag{0.19}$$

Hence,

$$\begin{aligned}\text{Curl}\,\mathbf{C} &= \nabla\times(\mathbf{A}\times\mathbf{B})\\ &= \mathbf{i}_x\left\{\frac{\partial}{\partial y}(A_xB_y - A_yB_x) - \frac{\partial}{\partial z}(A_zB_x - A_xB_z)\right\}\\ &\quad + \mathbf{i}_y\left\{\frac{\partial}{\partial z}(A_yB_z - A_zB_y) - \frac{\partial}{\partial x}(A_xB_y - A_yB_x)\right\}\\ &\quad + \mathbf{i}_z\left\{\frac{\partial}{\partial x}(A_zB_x - A_xB_z) - \frac{\partial}{\partial y}(A_yB_z - A_zB_y)\right\}\end{aligned}$$

Therefore, its x-component will be:

$$\begin{aligned}(\text{curl}\,\mathbf{C})_x &= \mathbf{i}_x\left[\left\{\left(\frac{\partial A_x}{\partial y}\right)B_y + A_x\left(\frac{\partial B_y}{\partial y}\right) - \left(\frac{\partial A_y}{\partial y}\right)B_x - A_y\left(\frac{\partial B_x}{\partial y}\right)\right\}\right.\\ &\quad \left. - \left\{\left(\frac{\partial A_z}{\partial z}\right)B_x + A_z\left(\frac{\partial B_x}{\partial z}\right) - \left(\frac{\partial A_x}{\partial z}\right)B_z - A_x\left(\frac{\partial B_z}{\partial z}\right)\right\}\right]\\ &= \mathbf{i}_x\left[A_x\left(\frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z}\right) - B_x\left(\frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z}\right)\right.\\ &\quad \left. + \left(B_y\frac{\partial A_x}{\partial y} + B_z\frac{\partial A_x}{\partial z}\right) - \left(A_y\frac{\partial B_x}{\partial y} + A_z\frac{\partial B_x}{\partial z}\right)\right]\\ &= \mathbf{i}_x\left\{A_x(\text{div}\,\mathbf{B}) - A_x\frac{\partial B_x}{\partial x}\right\} - \mathbf{i}_x\left\{B_x(\text{div}\,\mathbf{A}) - B_x\frac{\partial A_x}{\partial x}\right\} +\end{aligned}$$

$$\left\{\mathbf{i}_x(\mathbf{B}\cdot \text{grad } A_x) - \mathbf{i}_x B_x \frac{\partial A_x}{\partial x}\right\} - \left\{\mathbf{i}_x(\mathbf{A}\cdot \text{grad } B_x) - i_x A_x \frac{\partial B_x}{\partial x}\right\}$$

Writing in terms of the scalar directional derivative operators and cancelling out the +ve and –ve terms,

$$(\text{curl } \mathbf{C})_x = \mathbf{i}_x A_x(\text{div } \mathbf{B}) - \mathbf{i}_x B_x(\text{div } \mathbf{A}) + \mathbf{i}_x(\mathbf{B}\cdot \text{grad } A_x) - \mathbf{i}_x(\mathbf{A}\cdot \text{grad } B_x)$$

We can get similar expressions for the *y*- and *z*- components of curl **C**.

Hence adding all the three components,

$$\begin{aligned}\text{curl } \mathbf{C} &= \nabla \times (\mathbf{A} \times \mathbf{B}) \\ &= \mathbf{A} \text{ div } \mathbf{B} - \mathbf{B} \text{ div } \mathbf{A} + (\mathbf{B}\cdot \text{grad})\, \mathbf{A} - (\mathbf{A}\cdot \text{grad})\mathbf{B} \\ &= \mathbf{A}(\nabla\cdot\mathbf{B}) - \mathbf{B}(\nabla\cdot\mathbf{A}) + (\mathbf{B}\cdot\nabla)\mathbf{A} - (\mathbf{A}\cdot\nabla)\mathbf{B} \qquad (0.74b)\end{aligned}$$

0.8 TYPES OF VECTOR FIELDS

There is a certain amount of confusion and ambiguity in the minds of students when they talk of 'fields'. There is a tendency to assume implicitly that the 'field' is a 'region of space' which is quite wrong as would be seen from the following statement. *The electric field in the space enclosed by a charged hollow metal sphere is zero.* This is not to say that this space is zero. It is in fact a particular property of this space which is at zero level. Hence by the word field, we mean a particular property of the space (i.e. ensemble of measurements as a particular property).

0.8.1 Solenoidal and Irrotational Field (Lamellar)

Given, $\text{div } \mathbf{V} = 0 \quad \text{and} \quad \text{curl } \mathbf{V} = 0 \qquad (0.75)$

Since curl **V** = 0, this implies that **V** = grad *S* where *S* is the scalar potential.

$$\therefore \qquad \text{div grad } S = \nabla^2 S = 0 \qquad (0.76)$$

This equation is called the Laplace's equation, and such fields are called *Laplacian*. Examples of such fields are: electrostatic fields in free space, gravitational fields in free space, equilibrium thermal fields, magnetostatic fields in current-free region, static current field within a linear homogeneous isotropic conductor, and so forth.

0.8.2 Irrotational but not Solenoidal Field

In this case, $\text{curl } \mathbf{V} = 0 \quad \text{but} \quad \text{div } \mathbf{V} \neq 0 \qquad (0.77)$

Here again **V** = grad *S*, *S* being the scalar potential.

But $\text{div grad } S = \nabla^2 S \neq 0 \qquad (0.78)$

This equation is called the Poisson's equation and such fields are called Poissonian. Examples of such fields are: electrostatic fields in a charged medium, electrons in a thermionic tube, gravitational force inside a mass.

0.8.3 Solenoidal but not Irrotational Field

In this case, $$\text{div } \mathbf{V} = 0 \quad \text{but} \quad \text{curl } \mathbf{V} \neq 0 \tag{0.79}$$

Since $$\text{curl } \mathbf{V} \neq 0$$

$$\mathbf{V} = \text{curl } \mathbf{A}$$

where $\mathbf{A}$ is the vector potential. (0.80)

$$\therefore \quad \text{div } \mathbf{V} = \text{div curl } \mathbf{A} = 0 \quad \text{(This is a vector identity.)}$$

and

$$\text{curl } \mathbf{V} = \text{curl curl } \mathbf{A} \neq 0$$

$$= \text{grad div } \mathbf{A} - \nabla^2 \mathbf{A} \neq 0 \tag{0.81}$$

If, however, div $\mathbf{A} = 0$, then the above equation reduces to

$$\nabla^2 \mathbf{A} \neq 0 \tag{0.82}$$

This equation is similar to Poisson's equation, but in terms of a vector potential. Examples of such fields are: magnetic field within a conductor carrying a steady current, rotational motion of an incompressible fluid, time-varying electromagnetic field in a charge-free and current-free region.

0.8.4 Neither Irrotational Nor Solenoidal Field

This type of field satisfies the conditions:

$$\text{curl } \mathbf{V} \neq 0 \qquad \text{and} \qquad \text{div } \mathbf{V} \neq 0 \tag{0.83}$$

In this case, $$\mathbf{V} = \text{grad } S + \text{curl } \mathbf{A} \tag{0.84}$$

where S is the scalar potential and $\mathbf{A}$ is the vector potential.

$$\therefore \quad \text{div } \mathbf{V} = \text{div grad } S + \text{div curl } \mathbf{A} \neq 0$$

But div curl $\mathbf{A} = 0$, therefore, div grad $S \neq 0$. This is the Poisson's equation to determine S. Also

$$\text{curl } \mathbf{V} = \text{curl grad } S + \text{curl curl } \mathbf{A} \neq 0$$

But curl grad $S = 0$.

$$\therefore \quad \text{curl curl } \mathbf{A} \neq 0 \tag{0.85}$$

The relation (0.85) can be reduced to the form $\nabla^2 \mathbf{A} \neq 0$, by assuming div $\mathbf{A} = 0$, and then by solving for $\mathbf{A}$. This is the most generalized type of field. Such a general field vector can be decomposed into two fields, one being lamellar with the scalar potential S, and the other Solenoidal with the vector potential $\mathbf{A}$. This is sometimes known as 'Helmholtz' theorem. An example of such fields is that of rotational motion of compressible fluids.

0.8.4.1 A proof of the Helmholtz theorem

The Helmhotlz theorem is stated as: *A vector function is determined uniquely if the values of its curl and divergence are known at all points.*

Proof: Let $\mathbf{F}$ be a vector function, whose curl and div are known at all points, i.e.

$$\text{curl } \mathbf{F} = \nabla \times \mathbf{F} = \mathbf{P} \qquad \text{and} \qquad \text{div } \mathbf{F} = \nabla \cdot \mathbf{F} = Q \tag{0.86}$$

F cannot be determined as yet. So we assume that **F** is not determined uniquely by the Eq. (0.86), Hence, we can say

$$\nabla \times (\mathbf{F} + \mathbf{F}_0) = \mathbf{P} \qquad \text{and} \qquad \nabla \cdot (\mathbf{F} + \mathbf{F}_0) = Q \tag{0.87}$$

where $\mathbf{F}_0$ is a vector which is some function of position. As per the theorem, we will have to prove that $\mathbf{F}_0 = 0$.

From Eqs. (0.86) and (0.87), we have

$$\nabla \times \mathbf{F}_0 = 0 \qquad \text{and} \qquad \nabla \cdot \mathbf{F}_0 = 0 \tag{0.88}$$

We assume that $\nabla \cdot \mathbf{F}_0 = Q_0$ where Q_0 is an arbitrary function of position, though we know that $Q_0 = 0$. So instead of Eqs. (0.88), we have

$$\nabla \times \mathbf{F}_0 = 0 \qquad \text{and} \qquad \nabla \cdot \mathbf{F}_0 = Q_0 \tag{0.89}$$

with $Q_0 = 0$ in the final result. From the curl equation, we get

$$\mathbf{F}_0 = -\nabla V \tag{0.90}$$

where V is an arbitrary scalar function of position. Therefore, substituting in the second equation of (0.89), we get

$$\text{div grad } V = \nabla \cdot (\nabla V) = \nabla^2 V = -Q_0 \tag{0.91}$$

This is the Poisson's equation, whose solution is given by

$$V = \iiint_v \left(\frac{Q_0 \, dv}{4\pi r} \right) = \iiint_v \left\{ \frac{(\nabla \cdot \mathbf{F}_0) dv}{4\pi r} \right\}$$

The integration has to be performed over the whole space v. Now if $\nabla \cdot \mathbf{F}_0 = 0$ everywhere, then $V = 0$ over the whole space. Therefore, $\mathbf{F}_0 = 0$.

Thus, the vector **F** is uniquely defined, if its curl and divergence are known.

0.9 TIME VARIATION OF VECTORS

Vectors are quantities which are functions of space coordinates (magnitude and direction). However it is possible to superimpose on vectors, the functions of time as well. Particularly, since we shall be dealing with time-varying electromagnetic fields, the variation of these vectors (electric vectors such as electric field intensity, electric flux density, and current density, usually denoted by **E, D, and J** respectively; and magnetic vectors such as magnetic field intensity, magnetic flux density, usually denoted by **B** and **H** respectively) with the time, is an important parameter which has to be looked into. The time variations can be classified into two types: (a) any arbitrary time variation which can be of transient type, i.e. a sudden step function or delta function; (b) a periodic repeatable variation, i.e. a time-harmonic function, the most commonly used one in electromagnetism being the sinusoidal time-variation. This type of time-variation is used widely both in low frequency electrical machines as well as in high frequency electromagnetic waves and in radiation. So at present we will, for sometime, restrict our attention to 'sinusoidally time-varying vectors'.

0.9.1 Complex Representation of Time-harmonic Vectors

Let us suppose that the frequency of the source (which in our case would be time-varying charges and/or currents) producing the vector field (which would be the electric field **E** and the magnetic field **B**), is f. The product $2\pi f = \omega$ is usually referred to as the angular frequency. The (field) vectors will vary at all points with the same frequency. Thus such vectors can be written as

$$\mathbf{E}(x, y, z, t) = \mathbf{E}(x, y, z) \cos [\omega t + \theta_E (x, y, z)]$$
$$\mathbf{B}(x, y, z, t) = \mathbf{B}(x, y, z) \cos [\omega t + \theta_B(x, y, z)] \tag{0.92}$$

the vectors **E** and **B** being functions of space coordinates and time. The phases θ_E and θ_B are also functions of the coordinates and it should be noted that the field vectors are not necessarily at the same phase at all points at a given instant of time.

The formal analysis of such time-harmonic vectors becomes greatly simplified by a clever mathematical technique very often used in electrical engineering. The essence of this method is based on the following identities from the 'complex variable theory':

$$\cos x = \frac{\exp (j\omega t) + \exp (-j\omega t)}{2} \qquad \text{and} \qquad \sin x = \frac{\exp (j\omega t) - \exp (-j\omega t)}{2j}$$

for any x, where $j = \sqrt{-1}$.

Hence the vectors **E** and/or **B** can be written as

$$E(x, y, z, t) = E(x, y, z)\left[\frac{\exp \{j\omega t + \theta_E(x, y, z)\} + \exp \{-j\omega t + \theta_E(x, y, z)\}}{2}\right] \tag{0.93}$$

Also, we note that

$$\exp \left[\pm j(\omega t + \theta)\right] = \exp (\pm j\omega t) \exp (\pm j\theta)$$

where $\exp (\pm j\theta)$ is a function of space coordinates only. Hence if we have to write an equation of the form (choosing one of the Maxwell's field equations)

$$\text{curl } \mathbf{E} = \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{0.94}$$

it becomes, by using the complex variable notations,

$$\begin{aligned} &\exp (j\omega t) \nabla \times \left[E (x, y, z) \exp \left\{j\theta_E (x, y, z)\right\}\right] \\ &\quad + \exp (-j\omega t) \nabla \times \left[E (x, y, z) \exp \left\{-j\theta_E (x, y, z)\right\}\right] \\ &= -j\omega B (x, y, z) \exp \left\{j\theta_B (x, y, z)\right\} \exp (j\omega t) \\ &\quad + j\omega B (x, y, z) \exp \left\{-j\theta_B (x, y, z)\right\} \exp (-j\omega t) \end{aligned} \tag{0.95}$$

Since Eq. (0.95) must be satisfied for all values of t, we have to equate separately the coefficients of $\exp (j\omega t)$ and $\exp (-j\omega t)$, which gives us

$$\nabla \times \left[E (x, y, z) \exp \left\{j\theta_E (x, y, z)\right\}\right] = -j\omega \left[B (x, y, z) \exp \left\{j\theta_B (x, y, z)\right\}\right] \tag{0.96}$$

and

$$\nabla \times \left[E(x,y,z) \exp \left\{ -j\theta_E (x,y,z) \right\} \right] = j\omega \left[B(x,y,z) \exp \left\{ -j\theta_B (x,y,z) \right\} \right] \tag{0.97}$$

Thus, when the vectors vary with time, according to the simple harmonic law, both Eqs. (0.96) and (0.97) become equivalent to Eq. (0.94). It is more customary to use Eq. (0.96). The usual notation to denote a complex vector is to use a suffix c or a cap, i.e. $\mathbf{E}_c$ or $\hat{\mathbf{E}}$.

When the complex vectors are determined in a problem, their instantaneous values can be determined easily. For example, if the complex vector $\mathbf{E}_c(x, y, z)$ is known, its real value can be written in the form

$$\mathbf{E}_c(x, y, z) = \mathbf{E}(x, y, z) \exp \left[j \left\{ \omega t + \theta_E (x, y, z) \right\} \right] \tag{0.98}$$

where $\mathbf{E}(x, y, z)$ denotes a real function. Since

$$\exp (j\omega t) \exp \left[j\theta_E (x,y,z) \right] = \exp \left[j \left\{ \omega t + \theta_E (x, y, z) \right\} \right]$$

$$= \cos \{\omega t + \theta_E(x, y, z)\} + j \sin \{\omega t + \theta_E(x, y, z)\}$$

the correct value of $\mathbf{E}(x, y, z, t)$ is obtained by multiplying the equation by $\exp (j\omega t)$ and taking the real part of the resulting expression, i.e.

$$\mathbf{E}(x, y, z, t) = \text{Re}\{\mathbf{E}_c(x, y, z) \exp (j\omega t)\}$$

$$= \mathbf{E}(x, y, z) \cos \{\omega t + \theta_E (x, y, z)\} \tag{0.99}$$

{*Note:* Re stands for 'the real part of'.}

0.9.2 Complex Representation of Rotating Vectors

A certain amount of caution is required when using complex notation and the representation of time-harmonic (or sinusoidally varying with time) vectors. For example, when we are representing all the Cartesian components of the vectors (in our case we have mostly the electromagnetic vectors) oscillating with the same angular frequency ω, we can represent each component by a 'rotating vector' in the complex plane. This complex plane should not be confused with the two-dimensional Cartesian x-y plane, nor the complex plane rotation with the rotation of a two-dimensional vector in the Cartesian plane. We will try to illustrate this point, even at the cost of some duplication in our present discussion. At first we will consider the so-called rotation (or in fact oscillation) in the complex plane. We will also use the electric and the magnetic field vectors (i.e. **E** and **B**, respectively), though what we say here would be applicable in general to all vectors varying sinusoidally with time. Let us consider a particular location in space, i.e. the x-component of the (electric) vector, i.e. the real part or the reference component of the rotating vector in the complex plane which is given by

$$E_x = E_x^0 \exp (j\omega t), \quad \text{where} \quad \left| E_x^0 \right| = E_x^0 \exp \left(j \arg E_x^0 \right) \tag{0.100}$$

Similarly the other components of **E**, i.e. E_y and E_z and the magnetic field components of **B**, i.e. B_x, B_y and B_z can be represented on the complex plane (Figure 0.17).

It should be noted that Figure 0.17 shows not the actual components of the **E** and **B** vectors, but the real parts of the rotating vectors that give the actual electromagnetic vibrations. If the positions of the rotating vectors in Figure 0.17 at the time zero correspond to (E_x^0, E_y^0, E_z^0) and

(B_x^0, B_y^0, B_z^0), then these complex quantities give the amplitudes and the phases of the six vibrations, and constitute their 'complex amplitudes'. This rotation is thus described by a complex oscillation function, exp ($j\omega t$), with ω real or complex. This technique simplifies the operations of differentiation and integration with respect to time. Thus (a) differentiation with respect to time involves multiplication by $j\omega$ and (b) integration with respect to time reduces to division by $j\omega$.

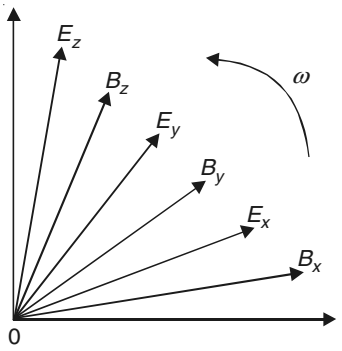


Figure 0.17 Complex plane representation of 'rotating vector'. The figure shows the reference component, or the real parts of the three-dimensional Cartesian components of the **E** and **B** vector at a particular point in space.

0.9.3 Magnitudes of Vectors in Complex Representation

During our discussion of sinusoidally time-varying vectors, for which is each Cartesian component oscillating with a specified angular frequency ω, we are involved in using in three-dimensional space the vectors for which each Cartesian component is a rotating vector in the complex plane (see Section 0.9.2). It is very important to differentiate a vector W in the complex plane, having real and quadrature components as U, V, respectively, expressed as

$$W = U + jV \tag{0.101}$$

from a vector **A** in the three-dimensional space, having Cartesian components A_x, A_y, A_z, expressed as

$$A = \mathbf{i}_x A_x + \mathbf{i}_y A_y + \mathbf{i}_z A_z \tag{0.102}$$

It is important to differentiate between the magnitude of a vector in the complex plane and the magnitude of a vector in the three-dimensional Cartesian space.

The magnitude of the two-dimensional complex plane vector is

$$|W| = U^2 + V^2 \tag{0.103}$$

with U, V being both real quantities.

The magnitude of the three-dimensional vector [Eq. (0.102)] will be expressed as A, where

$$A^2 = A_x^2 + A_y^2 + A_z^2 \tag{0.104}$$

in which each of the Cartesian components, A_x, A_y, A_z will be in general, complex and hence A is also complex. Therefore, although A in Eq. (0.104) is the magnitude of the three-dimensional vector given by Eq. (0.102), it is nevertheless a vector in the complex plane of magnitude A where

$$|A^2| = |A_x^2 + A_y^2 + A_z^2| \tag{0.105}$$

It must be clearly understood that the quantity on the right-hand side of Eq. (0.105), i.e. $\left|A_x^2 + A_y^2 + A_z^2\right|$ is not the same as $\left[\left|A_x^2\right| + \left|A_y^2\right| + \left|A_z^2\right|\right]$. This latter quantity is the sum of the squares of the magnitudes of the complex numbers A_x, A_y, A_z which constitute the Cartesian components of the three-dimensional vector **A** (each Cartesian component being complex), whereas the quantity $\left|A_x^2 + A_y^2 + A_z^2\right|$ is obtained by squaring each of the complex numbers A_x, A_y, A_z, which are then added together in the complex plane and then the magnitude of the complex number is evaluated.

A convenient way of obtaining $\left|A_x^2\right| + \left|A_y^2\right| + \left|A_z^2\right|$ is as follows: Starting with the 'complex conjugate' vector, which is also three-dimensional complex, i.e.

$$\mathbf{A}^* = \mathbf{i}_x A_x^* + \mathbf{i}_y A_y^* + \mathbf{i}_z A_z^* \tag{0.106}$$

we take the scalar product of vectors **A** and $\mathbf{A}^*$ as

$$\begin{aligned}\mathbf{A} \cdot \mathbf{A}^* &= A_x A_x^* + A_y A_y^* + A_z A_z^* \\ &= \left|A_x^2\right| + \left|A_y^2\right| + \left|A_z^2\right| \end{aligned} \tag{0.107}$$

It should be further noted that in general,

$$\mathbf{A}^2 \neq \mathbf{A} \cdot \mathbf{A}^* \tag{0.108}$$

When we are handling oscillatory electromagnetic fields, each Cartesian component of each oscillatory electromagnetic vector is represented in the complex plane as the reference component of a rotating vector. It is very important not to confuse between the three-dimensional vector algebra involving scalar and vector products with the two-dimensional vector algebra in the complex plane involving complex products and quotients.

0.9.4 Complex Representation of a Vector Rotating in Cartesian Plane

So far we have been discussing the rotation of the complex vectors in the complex plane. This behaviour can be looked at as either rotation or oscillation in that (complex) plane and must be clearly distinguished from the rotation in the Cartesian plane. So we shall now show how we can consider such 'physical' rotation. When we are using a Cartesian coordinate system, a vector **E** varying sinusoidally with time can be represented as (which we have already discussed earlier),

$$\begin{aligned}\mathbf{E} &= \mathbf{i}_x \,\mathrm{Re}[E_x \exp(j\omega t)] + \mathbf{i}_y \,\mathrm{Re}[E_y \exp(j\omega t)] + \mathbf{i}_z \,\mathrm{Re}[E_z \exp(j\omega t)] \\ &= \mathrm{Re}[(\mathbf{i}_x E_x + \mathbf{i}_y E_y + \mathbf{i}_z E_z) \exp(j\omega t)]\end{aligned}$$

where ω is the angular frequency variable, E_x, E_y, E_z are complex quantities, and Re indicates the operation of taking the real part. The time-dependence of **E** can also be expressed in a simpler and more compact form as

$$\mathbf{E} = \mathrm{Re}\left[\mathbf{E}_c \exp(j\omega t)\right] = \left[\frac{\mathbf{E}_c \exp(j\omega t) + \mathbf{E}_c^* \exp(-j\omega t)}{2}\right]$$

where $\mathbf{E}_c = \mathbf{i}_x E_x + \mathbf{i}_y E_y + \mathbf{i}_z E_z$, i.e. $\mathbf{E}_c$ is a complex vector in the sense that its components E_x, E_y, and E_z are complex quantities. Its complex conjugate $\mathbf{E}_c^*$ is also a complex vector whose

components are the conjugates of the components of $\mathbf{E}_c$. It should be carefully noted that, in general, the magnitude of **E** does not vary sinusoidally with time. It varies sinusoidally (with time) only in the special case in which the *direction* of **E** is time-invariant, i.e. when E_x, E_y, and E_z have the same phase angle. In fact the magnitude of **E** may be time-invariant, as in the following case of this two-dimensional vector **E** in which $\mathbf{E}_c = \mathbf{i}_x jA + \mathbf{i}_y A$.

In this case, the two non-vanishing components of **E** are 90° out of phase; the magnitude of **E** is at all times equal to A and its direction rotates in counterclockwise at the angular velocity ω, as shown in the Figure 0.18.

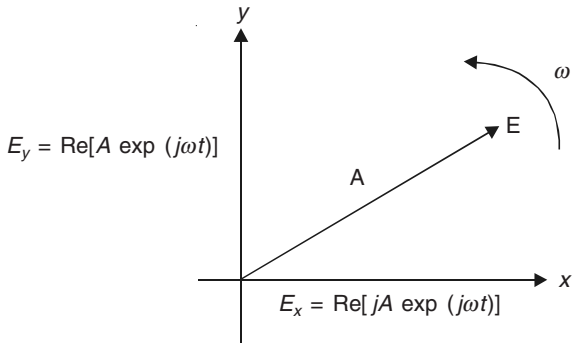


Figure 0.18 A two-dimensional rotating vector.

0.9.5 The Relationship between $\frac{d\mathbf{A}}{dt}$ and $\frac{\partial \mathbf{A}}{\partial t}$

So far in the previous sections (from 0.9.1 to 0.9.4) we have considered only the time-harmonic variation of the vectors, but now for the present discussion, the above-mentioned constraint does not exist.

For the present problem, there is a 'fixed' reference coordinate system (x-, y-, z- rectangular Cartesian system) and in this region there exists a vector field **A**, where

$$\mathbf{A} = \mathbf{i}_x A_x + \mathbf{i}_y A_y + \mathbf{i}_z A_z$$

For simplicity of understanding, we define 'a point in a moving body' in the region of the field under consideration. (However it must be clearly understood that the concept of the moving body in the field is not a necessary requirement for the present derivation, and the field at point external to the moving body would be equally valid for the analysis.) Now the time-rate of change of the vector **A** at a point (x, y, z) in the above-mentioned body can be evaluated in two different ways.

(i) The time-rate of change of **A** at a point fixed in the reference coordinate system, past which the body is moving, is denoted by $\frac{\partial \mathbf{A}}{\partial t}$—the partial derivative with respect to time with the values of x, y and z in the function $|\mathbf{A}| = f(x, y, z, t)$.

(ii) The time-rate of change of **A** at a point fixed in the body, and thus moving with it so that all the variables x, y, z and t can be changing, would be denoted by $\frac{d\mathbf{A}}{dt}$. This is then equivalent to the time-rate of change of **A** with reference to a coordinate system fixed in the moving body.

To find the relationship between $\frac{\partial \mathbf{A}}{\partial t}$ and $\frac{d\mathbf{A}}{dt}$ (refer to Figure 0.19), let a point fixed in the moving body have a velocity **v** which is given by

$$\mathbf{v} = \mathbf{i}_x v_x + \mathbf{i}_y v_y + \mathbf{i}_z v_z$$

Suppose, initially **A** has a constant value at such a point which is moving with the body, so that $\frac{d\mathbf{A}}{dt} = 0$.

But, in general, **A** varies with the position of the point in the body, and its value at the point P′ will be different from that at P (refer to Figure 0.19).

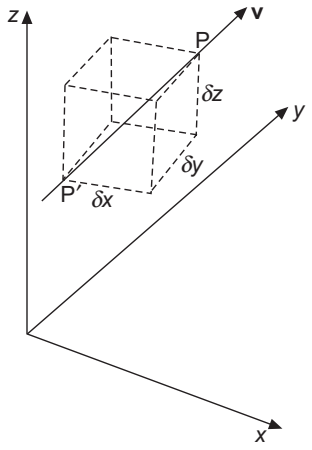


Figure 0.19 The points P and P′ to illustrate the relationship between $\frac{\partial \mathbf{A}}{\partial t}$ and $\frac{d\mathbf{A}}{dt}$.

Let the point P′ (in the body) move to the fixed point P in a time interval δt, so that

$$\mathrm{PP}' = \mathbf{v}\delta t = \mathbf{i}_x v_x \delta t + \mathbf{i}_y v_y \delta t + \mathbf{i}_z v_z \delta t$$
$$= \mathbf{i}_x \delta x + \mathbf{i}_y \delta y + \mathbf{i}_z \delta z$$

Let the value of the vector at the point P at the time-instant t be **A**. Then at the instant $t + \delta t$, it would have changed to the value **A′** which had existed at the point P′ at the time-instant t. Thus the x-component of **A** has changed by an amount δA_x which is given by:

$$\delta A_x = -\left\{ \frac{\partial A_x}{\partial x}\,\delta x + \frac{\partial A_x}{\partial y}\,\delta y + \frac{\partial A_x}{\partial z}\,\delta z \right\}$$

There will be similar expressions for the other two components of $\delta\mathbf{A}$, i.e. δA_y and δA_z. Hence adding all the three components vectorially, we get

$$\delta\mathbf{A} = \mathbf{i}_x \delta A_x + \mathbf{i}_y \delta A_y + \mathbf{i}_z \delta A_z$$

$$= -\mathbf{i}_x \left\{ \frac{\partial A_x}{\partial x}\delta x + \frac{\partial A_x}{\partial y}\delta y + \frac{\partial A_x}{\partial z}\delta z \right\}$$

$$-\mathbf{i}_y \left\{ \frac{\partial A_y}{\partial x}\delta x + \frac{\partial A_y}{\partial y}\delta y + \frac{\partial A_y}{\partial z}\delta z \right\}$$

$$-\mathbf{i}_z \left\{ \frac{\partial A_z}{\partial x}\delta x + \frac{\partial A_z}{\partial y}\delta y + \frac{\partial A_z}{\partial z}\delta z \right\}$$

$$= -(\mathbf{v}\cdot\nabla)\,\mathbf{A}$$

where $(\mathrm{v}\cdot\nabla)$ is the scalar directional derivative operator (refer to Section 0.5.1.1.).

So, if **A** is changing at the point fixed in the body, and moving with it at a rate $\frac{d\mathbf{A}}{dt}$, then in the limit as $\delta t \to 0$, this rate will be same at all the points on the infinitesimal path PP′.

Hence, when $\delta t \to 0$, the value of **A** at the point P at the instant $t + \delta t$ = the value of **A** existing at the point P′ at the instant t plus the changes in **A** at the moving point as it moves from P′ to P.

∴ The total time-rate of change at a point fixed in the reference coordinate system

$$= \frac{d\mathbf{A}}{dt} + \{-(\mathbf{v}\cdot\nabla)\mathbf{A}\}$$

i.e.
$$\frac{\partial \mathbf{A}}{\partial t} = \frac{d\mathbf{A}}{dt} - (\mathbf{v}\cdot\nabla)\mathbf{A}$$

which is the required relation.

So, we remind again that though the concept of 'the moving point being in the body' has been used in this analysis, the result is not restricted to points 'only in a moving material body.' This analysis applies equally to any point in the *external* field of a moving field-source and fixed in relation to the source. Thus the body described here includes both a field source and its field, be it external or internal.

This result has *nothing* to do with the theory of Relativity, since only *one* observer is involved. The vector **A** has been measured by the observer whose reference system is the coordinate axis system (*x*-, *y*-, *z*-) and he alone is calculating both $\frac{\partial \mathbf{A}}{\partial t}$ and $\frac{d\mathbf{A}}{dt}$.

The theory of Relativity deals with the value of **A** which is measured in a reference system moving with the body, and can be different from that measured in the system (*x*-, *y*-, *z*-).

Note:

1. When the velocity vector **v** is a constant, $(\mathbf{v}\cdot \mathrm{grad})\,\mathbf{A} = \mathbf{v}(\nabla\cdot\mathbf{A}) - \mathrm{curl}\,(\mathbf{v}\times\mathbf{A})$
   This has been done in Problem 0.16 of *Electromagnetism—Problems with Solutions*, 2nd Edition, 2008.
2. Also compare the result of this section (along with the above derivation of the note 1) with the derivation of generalization of the 'Faraday's law of Electromagnetic Induction' discussed in Sections 10.2 to 10.4 (pp. 313–322).

0.9.6 Conversion of a Vector from One Coordinate System to Another

We will consider briefly the interconvertibility of the vectors between the three most commonly used coordinate systems, though this conversion could also be written in terms of the general curvilinear coordinate system as well.

0.9.6.1 Conversion between rectangular Cartesian coordinate system and spherical polar coordinate system

We will first write down the expressions in terms of the unit vectors in these two coordinate systems though these expressions are quite general and hold for any vectors. First, we express the unit vectors of the spherical polar coordinate system in terms of the unit vectors in the rectangular Cartesian system as:

$$\mathbf{i}_\rho = \mathbf{i}_x \sin\theta\cos\phi + \mathbf{i}_y \sin\theta\sin\phi + \mathbf{i}_z \cos\theta$$

$$\mathbf{i}_\theta = \mathbf{i}_x \cos\theta\cos\phi + \mathbf{i}_y \cos\theta\sin\phi - \mathbf{i}_z \sin\theta$$

$$\mathbf{i}_\phi = -i_x \sin\phi + \mathbf{i}_y \cos\phi$$

Hence writing in compact matrix form for any vector (**A**, say)

$$\begin{bmatrix} \mathbf{A}_\rho \\ \mathbf{A}_\theta \\ \mathbf{A}_\phi \end{bmatrix} = \begin{bmatrix} \sin\theta\cos\phi & \sin\theta\sin\phi & \cos\theta \\ \cos\theta\cos\phi & \cos\theta\sin\phi & -\sin\theta \\ -\sin\phi & \cos\phi & 0 \end{bmatrix} \begin{bmatrix} \mathbf{A}_x \\ \mathbf{A}_y \\ \mathbf{A}_z \end{bmatrix}$$

Next expressing the unit vectors of the Cartesian system in terms of the spherical polar system, we get:

$$\mathbf{i}_x = \mathbf{i}_\rho \sin\theta\cos\phi + \mathbf{i}_\theta \cos\theta\cos\phi - \mathbf{i}_\phi \sin\phi$$

$$\mathbf{i}_y = \mathbf{i}_\rho \sin\theta\sin\phi + \mathbf{i}_\theta \cos\theta\sin\phi + \mathbf{i}_\phi \cos\phi$$

$$\mathbf{i}_z = \mathbf{i}_\rho \cos\theta - \mathbf{i}_\theta \sin\theta$$

Hence, writing in matrix form for any vector,

$$\begin{bmatrix} \mathbf{A}_x \\ \mathbf{A}_y \\ \mathbf{A}_z \end{bmatrix} = \begin{bmatrix} \sin\theta\cos\phi & \cos\theta\cos\phi & -\sin\phi \\ \sin\theta\sin\phi & \cos\theta\sin\phi & \cos\phi \\ \cos\theta & -\sin\theta & 0 \end{bmatrix} \begin{bmatrix} \mathbf{A}_\rho \\ \mathbf{A}_\theta \\ \mathbf{A}_\phi \end{bmatrix}$$

0.9.6.2 Conversion between rectangular Cartesian coordinate system and cylindrical polar coordinate system

Expressing the unit vectors of the cylindrical polar system, in terms of those in the rectangular system, we have

$$\mathbf{i}_r = \mathbf{i}_x \cos\phi + \mathbf{i}_y \sin\phi$$

$$\mathbf{i}_\phi = -\mathbf{i}_x \sin\phi + \mathbf{i}_y \cos\phi$$

$$\mathbf{i}_z = \qquad\qquad \mathbf{i}_z$$

Hence expressing in the matrix form,

$$\begin{bmatrix} \mathbf{A}_r \\ \mathbf{A}_\phi \\ \mathbf{A}_z \end{bmatrix} = \begin{bmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{A}_x \\ \mathbf{A}_y \\ \mathbf{A}_z \end{bmatrix}$$

Next, expressing the unit vectors of the rectangular coordinate system in terms of the cylindrical coordinate system,

$$\mathbf{i}_x = \mathbf{i}_r \cos\phi - \mathbf{i}_\phi \sin\phi$$

$$\mathbf{i}_y = \mathbf{i}_r \sin\phi + \mathbf{i}_\phi \cos\phi$$

$$\mathbf{i}_z = \qquad\qquad \mathbf{i}_z$$

Hence, for the general vector in terms of matrix equation

$$\begin{bmatrix} \mathbf{A}_x \\ \mathbf{A}_y \\ \mathbf{A}_z \end{bmatrix} = \begin{bmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{A}_r \\ \mathbf{A}_\phi \\ \mathbf{A}_z \end{bmatrix}$$

0.9.6.3 Conversion between cylindrical polar coordinate system and spherical polar coordinate system

Starting with the unit vectors of the spherical polar system,

$$\mathbf{i}_\rho = \mathbf{i}_r \sin\theta + \mathbf{i}_z \cos\theta$$

$$\mathbf{i}_\theta = \mathbf{i}_r \cos\theta - \mathbf{i}_z \sin\theta$$

$$\mathbf{i}_\phi = \mathbf{i}_\phi$$

Hence

$$\begin{bmatrix} \mathbf{A}_\rho \\ \mathbf{A}_\theta \\ \mathbf{A}_\phi \end{bmatrix} = \begin{bmatrix} \sin\theta & 0 & \cos\theta \\ \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{A}_r \\ \mathbf{A}_\phi \\ \mathbf{A}_z \end{bmatrix}$$

Expressing the unit vectors of the cylindrical polar system in terms of the spherical polar system, we have

$$\mathbf{i}_r = \mathbf{i}_\rho \sin\theta + \mathbf{i}_\theta \cos\theta$$

$$\mathbf{i}_\phi = \qquad \mathbf{i}_\phi$$

$$\mathbf{i}_z = \mathbf{i}_\rho \cos\theta - \mathbf{i}_\theta \sin\theta$$

Hence

$$\begin{bmatrix} \mathbf{A}_r \\ \mathbf{A}_\phi \\ \mathbf{A}_z \end{bmatrix} = \begin{bmatrix} \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \\ \cos\theta & -\sin\theta & 0 \end{bmatrix} \begin{bmatrix} \mathbf{A}_\rho \\ \mathbf{A}_\theta \\ \mathbf{A}_\phi \end{bmatrix}$$

0.9.6.4 General comments

It should be noted that only in the rectangular Cartesian system, all the three unit vectors (parallel to the coordinate axes) are equal in magnitude and have 'fixed' directions. In the cylindrical coordinate system, only $\mathbf{i}_z$ has fixed direction whilst both $\mathbf{i}_r$ and $\mathbf{i}_\phi$ keep on changing their directions, However, all the three unit vectors in the spherical coordinate system keep on changing their directions as they move from one point to another. Hence the derivatives of $\mathbf{i}_r$, $\mathbf{i}_\phi$, $\mathbf{i}_\rho$, $\mathbf{i}_\theta$, $\mathbf{i}_\phi$ with respect to coordinate directions are non-zeroes, as shown in Appendix 4, Section A.4.14.

PROBLEMS

0.1 Under what circumstances are the vectors $\mathbf{A} \times (\mathbf{B} \times \mathbf{C})$ and $\mathbf{B} \times (\mathbf{A} \times \mathbf{C})$ equal?

0.2 If the vector $\mathbf{A}$ has constant magnitude, show that the vectors $\mathbf{A}$ and $\dfrac{d\mathbf{A}}{dt}$ are perpendicular, provided $\dfrac{d\mathbf{A}}{dt} \neq 0$.

0.3 If V is a scalar function of x and y, show that the divergence of the vector field $\mathbf{F} = \mathbf{i}_z \times \text{grad } V$ is always zero.

0.4 The direction of a vector $\mathbf{A}$ is radially outwards from the origin, and its magnitude is kr^n, where

$$r^2 = x^2 + y^2 + z^2$$

Find the value of n for which div $\mathbf{A} = 0$.

Ans.: $n = -2$

0.5 A vector $\mathbf{A}$ is given in cylindrical coordinates as

$$\mathbf{A} = \mathbf{i}_r\, 2 \cos\phi + \mathbf{i}_\phi r$$

Evaluate the line integral of $\mathbf{A}$ around the contour in the $z = 0$ plane bounded by $+x$- and $+y$-axes and the arc of the circle of radius 1 unit. Check the answer by performing the appropriate surface integral of curl $\mathbf{A}$.

0.6 A point P moves so that its position vector $\mathbf{r}$ relative to another point O satisfies the equation

$$\frac{d\mathbf{r}}{dt} = \boldsymbol{\omega} \times \mathbf{r}$$

where $\boldsymbol{\omega}$ is a constant vector. Prove that P describes a circle with constant velocity.

1 The Electrostatic Field in Free Space (in Absence of Dielectrics)

1.1 INTRODUCTION

The science of electromagnetism deals with mutual interactions of electric charges, both at rest as well as in motion. The branch of electromagnetism which deals with the interactions of the charges at rest is called *electrostatics*. In this text we will deal mostly with large aggregates of elemental electric charges. Such elemental charges are the building blocks of atoms of all substances, i.e. the electrons which are negatively charged and rotate round the nucleus which contains positively charged protons and neutrons which are electrically neutral. Normally the atoms of a body are electrically neutral, the positive charges of the protons being equal to the total negative charges of the electrons. When a body is charged, some of its atoms are taken out of their neutral condition by either adding or subtracting one or more electrons. The atom is then said to be ionized, positively if there is a deficiency of electrons, and negatively if there is an excess of electrons.

Since here we are going to concentrate on the interactions between large ensembles of these charges (i.e. atoms), we shall not consider the interactions between the charged particles inside the individual atoms—a topic which is the subject matter of the theories of microscopic properties of matter. We restrict ourselves to only the macroscopic aspect of the phenomenon. Since the number of charged particles in any volume to be considered is going to be quite large, the concept of infinitesimally small volume in the rigorous mathematical sense is not permitted. A limit has to be imposed on the size of the 'infinitesimally small volume'. It has to be very small, but still large enough to contain a large number of elemental charges. Such a small volume is known as 'physically small volume' and its length dimensions are much larger than the sub-atomic distances. It is not possible to give a unique order of magnitude for the physically small volume, but a volume of 10^{-6} cubic centimetres can be considered as an upper limit. Since roughly speaking a cubic centimetre of a rigid body contains between 10^{22} and 10^{24} atoms*, the specified small volume would contain some 10^{16} to 10^{18} atoms. Although the exact number of atoms for different substances would vary greatly, the upper limit mentioned above for the physically small volume is a reasonable one.

*The basis for this limit is the Avogadro constant which has the value 6.2252×10^{23} per mole. Refer to *Physics* by Resnik & Halliday, Vol. 1, p. 550.

Even though from the macroscopic point of view, physical quantities inside any 'physically small volume' can vary significantly from one point to another, such variations are of no importance when considering interactions between large aggregates of electrical charges. It is sufficient to know the average values of the quantities in a physically small volume. Such an approach is known as the *macroscopic* approach, and the averaged physical quantities are also called macroscopic. The present text will deal with macroscopic phenomena, but to understand and to explain the mechanism of certain interactions, it may be necessary at times to consider the model of the atomic structure of matter.

We shall start the present chapter by discussing the interactions between the electric charges in free space and subsequently introduce the effects of the material media of different kinds. One type of substances is known as *dielectrics* or *insulators* in which there are practically no freely movable charges in their structures. Substances having only a relatively small number of free charges in their structures are known as *semiconductors*. Those substances in which there are elemental electric charges, practically free to move inside them, are known as *conductors*. Examples of this type are metals (in which the electrons in the outer shell are very loosely attached to the rest of the atom) and electrolytic solutions (with movable charges called ions).

1.2 THE LAW OF FORCE BETWEEN CHARGED PARTICLES (COULOMB'S LAW)

The experimenters of the eighteenth century knew that the objects could be charged with electricity so that they would attract or repel each other. Comparing these forces with the gravitational attraction, these pioneers must have argued about a similar law for the electrostatic forces. It was a French engineer, Charles A. Coulomb who announced an experimental verification of this law in 1785. The law was enunciated as follows:

> *The force between two charged particles at rest is proportional to the product of the charges and inversely proportional to the square of the distance between them.*

The force acts along the line joining the particles, and is a repulsion if the charges are of the same sign, and an attraction if their signs are opposite. This law holds good when the charges are in vacuo, or when the particles are surrounded by air, or any other gas, or by a large class of liquids or solids known as dielectrics. It should be however noted that the extension of this law to particles embedded in solids is a generalization from experience, for it is impossible to measure directly the force in such a case. The surrounding medium, whilst it affects the magnitude of the force, does not affect the form of the law. Hence the force F between the particles bearing the charges Q_1 and Q_2 and separated by a distance r, is given by

$$F = \frac{\lambda Q_1 Q_2}{r^2} \tag{1.1}$$

where the multiplier λ depends on the surrounding medium. The law assumes that the surrounding medium is uniform not only between the charges, but also to a distance (in all directions) great compared with r. The precise value of λ depends on this assumption. Since all charges are built from protons or electrons which are elementary particles of equal charge, the charge of the proton or the electron can be considered as the natural unit for charges. However the unit used in practice

is 'coulomb' which is equal to 6.242×10^{18} times the electron charge, and this unit is connected with the basic units of mass, length, and time. The force between the particles bearing the charges Q_1, Q_2 coulombs, and separated by a distance r metres, is given in vacuo by

$$F = \frac{(8.988 \times 10^9)\, Q_1 Q_2}{r^2} \text{ newtons} \tag{1.2}$$

In air at normal temperature and pressure, this force is less by only 0.06 per cent and hence for all practical purposes, vacuum and air need not be differentiated. (In this context vacuum is often referred to as free space.) In the SI system of units, this constant is expressed as $\lambda = 1/4\pi\varepsilon_0$, where ε_0 is called the permittivity of free space (or absolute permittivity) and has the value

$$\varepsilon_0 = 8.854 \times 10^{-12} = \frac{1}{36\pi \times 10^9} \text{ coulomb}^2\text{/newton-m}^2 \text{ (or farads/metre—to be proved later)} \tag{1.3}$$

As coulomb is a large quantity, comparing it with the c.g.s. electrostatic unit, we have

$$1 \text{ coulomb} = 2.998 \times 10^9 \text{ electrostatic units}$$

So it will be seen that it is convenient to approximate coulomb by 3×10^9 e.s. units and λ by 9×10^9, so that these become easy to memorize.

The force F expressed as a vector can be written as

$$\mathbf{F}_{12} = \frac{1}{4\pi\varepsilon_0}\left(\frac{Q_1 Q_2}{r^2}\right)\mathbf{u}_{12} \tag{1.4}$$

where $\mathbf{u}_{12}$ is the unit vector directed from the charge Q_1 to Q_2, and $\mathbf{F}_{12}$ is the vector force by which the charge Q_1 acts on the charge Q_2 (Figure 1.1).

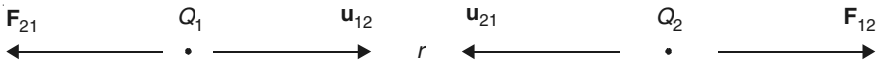


Figure 1.1 The two point charges and the reference directions of the vectors entering in the Coulomb's law.

In practical terms, the electrostatic forces are much larger than the gravitational forces. For example, the gravitational attraction between particles of masses m_1, m_2 kilograms separated by r metres is

$$F_g = \frac{(6.67 \times 10^{-11})\, m_1 m_2}{r^2} \text{ N} \tag{1.5}$$

which is equal to the electrostatic force between the charges of 0.86×10^{-10} C at the same spacing.

1.3 THE PRINCIPLE OF SUPERPOSITION

When there are more than two point charges, the total force on any one of them is the vector sum of the forces on it due to each charge considered separately. This is known as the 'principle of superposition'. This is again an experimental fact.

Thus if there are n point charges Q_1, Q_2, Q_3, …, Q_n distributed in a region, then the force on the charge Q at a point A is given by (Figure 1.2).

$$\mathbf{F}_{\mathrm{A}} = \sum_{k=1}^{n} \mathbf{r}_{k\mathrm{A}} \frac{QQ_k}{4\pi\varepsilon_0 r_k^2} \tag{1.6}$$

where r_k is the distance between Q_k and Q, and $\mathbf{r}_{k\mathrm{A}}$ is the unit vector in the direction of r_k.

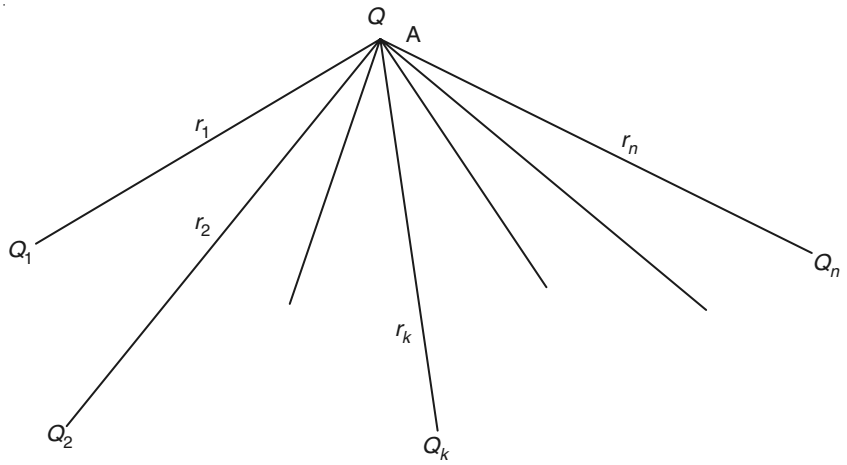


Figure 1.2 Distributed charges.

1.4 THE ELECTRIC FORCE (PER UNIT CHARGE) AND THE CONCEPT OF ELECTRIC FIELD

The electric force at a point is defined as the force experienced at that point by a particle bearing a unit positive charge placed at that specified point. It is denoted by the vector **E** and its magnitude by E. The region of the space where such a force exists is the region of the electric field, and the vector **E** is called the 'electric field intensity'.

To start with, in the region of interest, if there are two point charges (as shown in Figure 1.1), the force at the location of the point charges is given by Eq. (1.4). Now let the charge Q_2 be 1 C, then

$$\text{the force at that point} = \frac{Q_1}{4\pi\varepsilon_0 r^2} \tag{1.7}$$

This is defined as the electric field at that point, that is, **E** (the electric field intensity). It is a vector in the same direction as $\mathbf{F}_{12}$.

$$\therefore \qquad \mathbf{E} = \left(\frac{Q_1}{4\pi\varepsilon_0 r_k^2} \right) \mathbf{r}_1 \tag{1.8}$$

where $\mathbf{r}_1$ is the unit vector in the direction of the line joining Q_1 with the unit charge at the location of Q_2. Also,

$$\mathbf{F}_{12} = Q_2\mathbf{E} \tag{1.9}$$

The unit of **E** is newton/coulomb, and this will later be shown to be volts/metre.

If there are more than two point charges in the region, that is, n point charges as shown in Figure 1.2, then the electric field intensity at the point A is given by

$$\mathbf{E}_{\mathrm{A}} = \sum_{k=1}^{n} \mathbf{r}_{k1} E_{k\mathrm{A}} \tag{1.10}$$

where

$$E_{k\mathrm{A}} = \frac{Q_k}{4\pi\varepsilon_0 r_k^2}$$

To visualize mentally the electric field, we use the concept of the 'lines of force'. A line of force is a continuous curve, whose direction at every point is that of the vector **E**. Hence the lines of force from a single point charge would be radial straight lines extending outwards. Lines of force for point charges are shown in Figure 1.3. The whole space of the field is imagined to be filled with such lines and a charge brought into the field would experience a force whose direction is that of the line of force at that point where the extraneous charge is brought in. Here we meet with a conceptual difficulty. It should be noted that this intruding charge is not a passive device. It is by itself a source of electric force and the existing lines of force in its vicinity would get distorted by its presence. And yet when we say that the force on this extraneous charge is QE, the value of E to be used is that value which the electric force would have if the intruding charge were not there. The point we should note is that as a method of calculating the force on the charged particle, this works correctly, but as a picture of what is happening in the electric field it is unsatisfactory.

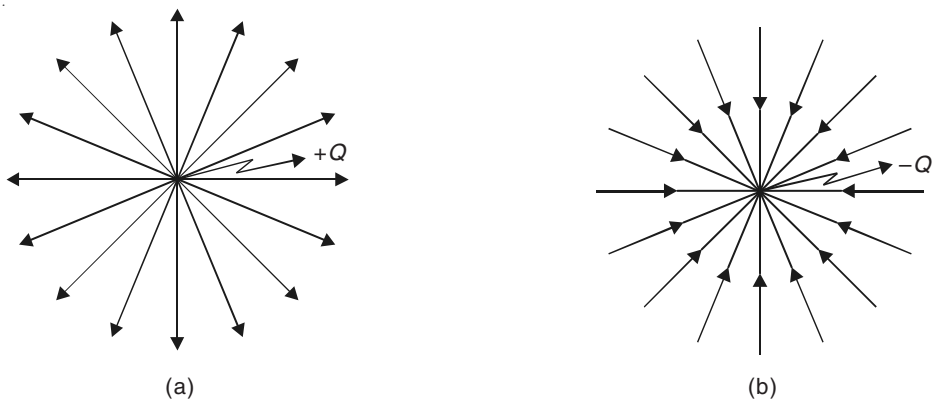


Figure 1.3 Graphical representation of electric lines of force of a point charge (a) $+Q$, (b) $-Q$, each in isolation.

1.5 THE ELECTRIC FIELD OF CONTINUOUS SPACE DISTRIBUTION OF CHARGES (GAUSS' THEOREM)

We started our discussion of electrostatic forces with two discrete point charges and extended our arguments to the space distribution of n point charges. The next logical step would be now to consider 'continuous space distribution of charges', which of course is the basis for the

macroscopic theory as applicable to a large number of problems in engineering. In such an approach, the macroscopic properties of these charges become important and so the volume charge density ρ_C has to be defined as

$$\rho_C = \frac{\text{sum of all charges in a physically small volume } \delta v}{\text{the enclosing volume element } \delta v}$$

Considering the whole region v occupied by the charges (totalling to Q), let the region be divided into similar volume elements. Then from Eq. (1.10) the electric field intensity can be written as

$$\mathbf{E} = \sum_v \left(\frac{\rho \, \delta v}{4\pi\varepsilon_0 r^2} \right) \mathbf{r}_1 \tag{1.11}$$

As the element δv tends to become smaller and smaller, the summation sign can be replaced by the volume integral over the region v, i.e.

$$\mathbf{E} = \left\{ \frac{1}{4\pi\varepsilon_0} \right\} \left[\iiint_v \left(\frac{\rho \, dv}{r^2} \right) \right] \mathbf{r}_1 \tag{1.12}$$

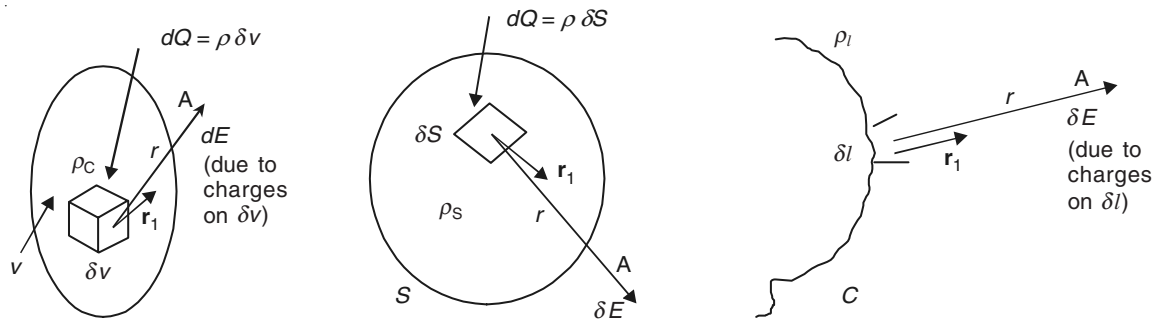
By similar arguments, the field intensity due to a surface distribution of charges ρ_S over a surface S can be shown to be

$$\mathbf{E} = \left\{ \frac{1}{4\pi\varepsilon_0} \right\} \left[\iint_S \left\{ \frac{\rho_S \, dS}{r^2} \right\} \right] \mathbf{r}_1 \tag{1.13}$$

and due to a line element of charges,

$$\mathbf{E} = \left\{ \frac{1}{4\pi\varepsilon_0} \right\} \left[\int_l \left\{ \frac{\rho_\ell \, dl}{r^2} \right\} \right] \mathbf{r}_1 \tag{1.14}$$

These three distributions are shown pictorially in Figure 1.4.



(a) Volume distribution of charges (b) Surface distribution of charges (c) Line distribution of charges

Figure 1.4 Distribution of charges.

1.5.1 The Flux of E Across a Surface

Before we consider Gauss' theorem as applied to the electric field intensity **E**, we look at the concept of 'the flux of **E** across a surface'. Let a surface S be divided into a meshwork of small elements, of which a typical one is δS. Even though S may be curved, the element δS is so small that it can justifiably be assumed to be flat and the line PN to be normal to δS. The line PM is the direction of **E** (the electric force) at the point P, and it makes an angle θ with the normal PN. Then ($E \cos\theta\, \delta S$) is the flux of E across the surface δS; and the sum of all the fluxes across the surfaces δS (i.e. $\Sigma\ E \cos\theta\, \delta S$) is the flux of E across the surface S. As the surface elements tend to become smaller and smaller, this quantity becomes

$$\iint_S E_n\, dS \quad \text{or} \quad \iint_S \mathbf{E} \cdot d\mathbf{S}$$

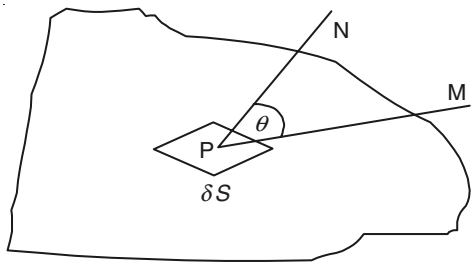


Figure 1.5 Surface *S*.

To make a physical analogy, when the integral is taken over a closed surface, the result equates the net outflow of the vector across the specified volume, for example, if the vector represents the velocity of a fluid, then the value of the integral is the net outflow of the fluid across the volume.

1.5.2 Gauss' Theorem

In general terms, the Gauss' theorem states that:

> *The flux of a vector quantity over any arbitrary closed surface is equal to (or proportional to) the strength of the enclosed sources of the vector.*

In the context of the electric field intensity and the electric charge, let us assume that a total charge Q is distributed over a volume v. Therefore, the charge density ρ is

$$\rho = \frac{Q}{v}$$

Also, let Σ be a closed surface enclosing this volume v, and P be a point on this surface. The next step is to evaluate the integral

$$\oiint_\Sigma \mathbf{E} \cdot d\mathbf{S}$$

over this surface Σ.

Consider an infinitesimally small element of charge $\rho\, \delta v$ ($= Q_k$) at the point O as shown in Figure 1.6. Then,

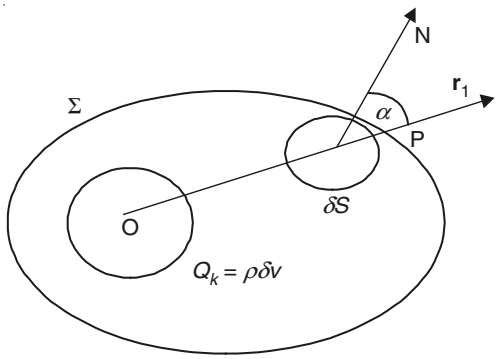


Figure 1.6 Construction for Gauss' theorem.

$$\mathbf{E} \cdot d\mathbf{S} = \frac{Q_k \mathbf{r}_1 \cdot d\mathbf{S}}{4\pi\varepsilon_0 r^2} = \frac{Q_k \, dS \cos\alpha}{4\pi\varepsilon_0 r^2} \tag{1.15}$$

where

r = the distance between the elemental charge Q_k in v and the point P on the closed surface Σ

$\mathbf{r}_1$ = the unit vector in the direction of r

α = the angle between $\mathbf{r}$ and $\delta\mathbf{S}$.

Now,

$$\frac{\delta S \cos\alpha}{r^2} = \text{solid angle subtended by } Q_k \text{ at } \delta S \text{ on } \Sigma.$$

= surface of the portion of the sphere of unit radius with its centre at O, cut by the cone subtended by δS with the vertex at O. (1.15a)

Integrating over the whole surface Σ, the integral = $Q_k \cdot 4\pi \cdot 1^2$, with $(4\pi \cdot 1^2)$ being the total solid angle subtended by Q_k over the whole surface Σ. Then by using the principle of superposition, we can consider all the charges enclosed by the closed surface Σ (which can be a space distribution of charges, in this case totalling to Q).

$$\therefore \quad \oiint_{\Sigma} \mathbf{E} \cdot d\mathbf{S} = \frac{Q}{\varepsilon_0} \tag{1.16}$$

Thus the Gauss' theorem gives a relation between the electric field and the charges in the system, which depends on the characteristic property of the medium, i.e. ε_0 in this case.

It should be noted that the shape of the volume considered is no restriction on this theorem, or if there are charges outside the volume under consideration. If the shape of the volume is such that the elemental cone cuts the closed surface more than once, say, three times, or any odd number of times (when the point O is inside the volume), then also Eq. (1.16) holds. If the charges are outside Σ, then the flux of the electric field intensity due to such charges would be zero.

Since Gauss' theorem gives a relationship between the flux of the electric field intensity **E**

and the enclosed charge Q, which depends on the characteristic property of the medium, we introduce a new vector **D** called the 'electric flux density' vector, such that, in vacuo,

$$\mathbf{D} = \varepsilon_0 \mathbf{E} \tag{1.17}$$

so that by Gauss' theorem

$$\oiint_{\Sigma} \mathbf{D} \cdot d\mathbf{S} = Q \tag{1.18}$$

where Q is the enclosed charge in the system. The reason for the choice of the name of **D** becomes obvious from Eq. (1.18).

Thus, so far we have derived the vector **E** starting from point charges, and subsequently using Gauss' theorem, we have extended these ideas to space distributions of charges, and hence the 'field concept' and in the process an auxiliary vector **D** has now been introduced. So essentially any problem dealing with the evaluation of **E** or **D** for a system of charges, becomes now a matter of dreaming up a suitable 'Gaussian Surface', which is defined as a convenient surface to which Gauss' theorem can be applied.

1.5.3 An Alternative Proof of Gauss' Theorem in Differential Form

Equation (1.18) can be expressed in a different form if we imagine the closed surface Σ to be the surface of a rectangular brick which may be very small. Figure 1.7 shows a brick with sides δx, δy, δz parallel to the coordinate axes of a right-handed coordinate system.

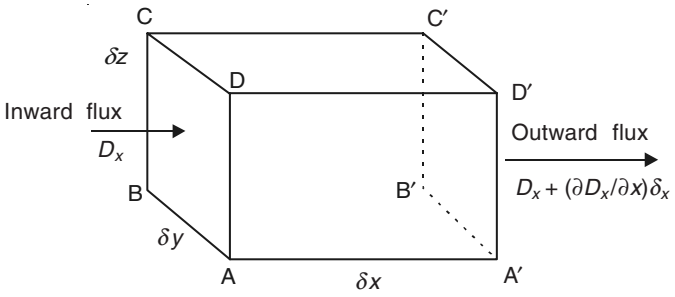


Figure 1.7 Elemental volume for Gauss' theorem in the rectangular coordinate system.

If there is a charge cloud of density ρ_C (C/m^3) in this region, then the charge within Σ is $\rho_C \delta x \delta y \delta z$.

To calculate the flux through the surface, let us consider first the parallel faces ABCD and A′B′C′D′. Their areas are $\delta y \delta z$ each, and so they give:

$$(\mathrm{ABCD}) \rightarrow \text{Inward flux} = D_x \delta y \delta z$$

$$(\mathrm{A'B'C'D'}) \rightarrow \text{Outward flux} = D_x + \left(\frac{\partial D_x}{\partial x}\right) \delta x \, \delta y \, \delta z$$

Therefore, the net outward flux from the two faces = $(\partial D_x/\partial x)\delta x \delta y \delta z$. And the other two sets of faces give similar expressions as above.

Thus, the Gauss' theorem takes the form:

$$\left(\frac{\partial D_x}{\partial x}+\frac{\partial D_y}{\partial y}+\frac{\partial D_z}{\partial z}\right)\delta x\,\delta y\,\delta z=\rho_C\delta x\,\delta y\,\delta z$$

or

$$\frac{\partial D_x}{\partial x}+\frac{\partial D_y}{\partial y}+\frac{\partial D_z}{\partial z}=\rho_C \tag{1.19}$$

The left-hand side of Eq. (1.19) is, $\lim_{\delta x\delta y\delta z\to 0}\dfrac{\text{net outward flux of } D \text{ from } \Sigma}{\text{volume enclosed by } \Sigma}$

The above quantity, a scalar, derived from the vector **D** is known as the divergence of **D** (i.e. div **D**). In general, therefore,

$$\text{div } \mathbf{D}=\lim_{\delta v\to 0}\frac{\oiint_\Sigma D_n\,dS}{\delta v} \tag{1.20}$$

where δv is the volume within Σ. Equation (1.19) shows that in Cartesian coordinates

$$\text{div } \mathbf{D}=\frac{\partial D_x}{\partial x}+\frac{\partial D_y}{\partial y}+\frac{\partial D_z}{\partial z} \tag{1.21}$$

but the importance of the general integral Eq. (1.20) is that it is not confined to the Cartesian coordinates. Other coordinates systems will be used later.

Now, Eq. (1.19) can be restated in the form

$$\text{div } \mathbf{D}=\rho_C \tag{1.22}$$

which does not imply any particular coordinate system. (This is also a restricted form of one of the Maxwell's equations.)

1.6 ELECTRIC POTENTIAL (OR ELECTROSTATIC POTENTIAL)

To facilitate the calculation of electric fields, the concept of electric potential is introduced, which is defined as follows:

The work done by the extraneous forces against the forces of the field in moving a unit charge from a point A *to a point* B *is called the increase of the electric potential between* A *and* B *(i.e. the potential difference between* A *and* B*).*

Using the vector notations, the electric potential is defined as

$$-\int_A^B \mathbf{E}\cdot d\mathbf{l}$$

PHYSICAL INTERPRETATION OF ELECTRIC POTENTIAL

$\mathbf{E}$ = force on a unit charge placed at the point under consideration
$d\mathbf{l}$ = vector length of the element where the charge is placed

then,

$$\mathbf{E} \cdot d\mathbf{l} = E\ dl\ \cos\ \alpha$$

where α is the angle between the directions $\mathbf{E}$ and $d\mathbf{l}$. That is,

$$\begin{aligned}\mathbf{E} \cdot d\mathbf{l} &= (E \cos \alpha)\ dl \\ &= \text{(component of the force in the direction of } dl) \times \text{length } dl \\ &= \text{work done in moving the unit charge over the distance } dl\end{aligned}$$

$$\therefore \int_{\mathbf{A}}^{\mathbf{B}} \mathbf{E} \cdot d\mathbf{l} = \text{work done in moving a unit charge from the point A to the point B} \quad (1.23)$$

Thus, this is a measure of the potential energy in the system, i.e. the energy distribution in the system. Its unit is:

Work done is in joules

Charge is in coulombs

$$\therefore \quad \text{Potential Difference (P.D.)} = \frac{\text{joules (J)}}{\text{coulombs (C)}} = \text{volts (V)}$$

A proof for the *work done* is independent of the path chosen for the line integral, and depends only on the terminal points. Let us consider the field due to a single point charge Q (Figure 1.8).

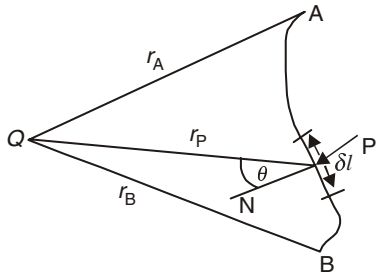


Figure 1.8 Potential due to a single point charge.

$\mathbf{E}$ at the point P on the path is given by

$$\left(\frac{Q}{4\pi\varepsilon_o r_P^2}\right)\mathbf{r}_{P1}$$

where $\mathbf{r}_{P1}$ is the unit vector in the direction of r_P.

$$\therefore \quad \mathbf{E} \cdot d\mathbf{l} = \left(\frac{Q}{4\pi\varepsilon_0}\right)\left(\frac{\mathbf{r}_{P1} \cdot d\mathbf{l}}{r_P^2}\right)$$

$$= \left(\frac{Q}{4\pi\varepsilon_0}\right)\left(\frac{dl \cos\theta}{r_P^2}\right)$$

$$= \left(\frac{Q}{4\pi\varepsilon_0}\right)\left(\frac{\delta r}{r_P^2}\right)$$

$$\therefore \quad -\int_{r_A}^{r_B} \mathbf{E} \cdot d\mathbf{l} = -\frac{Q}{4\pi\varepsilon_0} \int_{r_A}^{r_B} \frac{dr}{r_P^2}$$

$$= \frac{Q}{4\pi\varepsilon_0} \left(\frac{1}{r_B} - \frac{1}{r_A} \right) \tag{1.24}$$

Thus the work done is independent of the path chosen for the line integral. This can be extended to more complex charge distribution systems. Therefore, the expression

$$-\int_A^B \mathbf{E} \cdot dl$$

is a measure of the energy distribution in the field, i.e. a measure of the variations in the level of the potential energy at different points in the field. Separating the idea of the 'variation of level' from its physical basis of energy distribution, 'the variation of level' can be used as a scalar field associated with the original vector force field **E**.

Therefore, the potential difference between the points A and B is defined as

$$V_{AB} = -\int_A^B \mathbf{E} \cdot d\mathbf{l} \tag{1.25}$$

It should, however, be noted that this is not a definition of the absolute potential of a point. However, we can extrapolate from this definition to find the absolute potential. If the starting reference point is moved to infinity which is taken as the zero level (Figure 1.9), then this defines the absolute potential of a point, i.e.

$$V_A = -\int_\infty^A \mathbf{E} \cdot dl, \qquad V_B = -\int_\infty^B \mathbf{E} \cdot dl \tag{1.26}$$

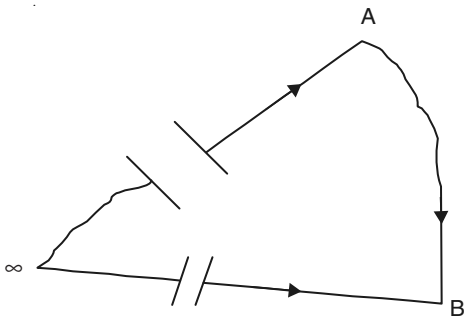


Figure 1.9 Absolute potential in terms of the potential of the point at infinity.

This definition agrees with the previously given definition of the potential difference (P.D.), i.e.

$$V_{AB} = -\int_{A}^{B} \mathbf{E} \cdot d\mathbf{l} = -\int_{\infty}^{B} \mathbf{E} \cdot d\mathbf{l} + \int_{\infty}^{A} \mathbf{E} \cdot d\mathbf{l} = V_B - V_A \tag{1.27}$$

This follows from the fact that the line integral of **E** around a closed path is zero, i.e. $\infty \rightarrow A \rightarrow B \rightarrow \infty$ is zero; a conclusion which was also implicit in the result derived in Eq. (1.24). Emphasizing this point, it follows that since the work done in conveying a charged particle from one point to another is independent of the path traversed, the work done in traversing a closed path would be zero (i.e. the line integral of **E** along a closed path is zero). Thus,

$$\oint_{C} \mathbf{E} \cdot d\mathbf{l} = 0 \tag{1.28}$$

Or referring back to the vector operators, it follows that

$$\text{curl } \mathbf{E} = \nabla \times \mathbf{E} = 0 \tag{1.29}$$

This is the consequence of the conservation of energy in this type of field. Such a field, which satisfies the above equation is called a 'conservative field'.

1.6.1 Electric Field and Electric Potential

It is seen that when the electric field **E** in a region is given, the potential field can be found; or vice versa if the potential field V is known, how do we find the **E** field?

By the definition of potential difference, over a small length δl if the potential drop is δV, then

$$\delta V = -\mathbf{E} \cdot \delta \mathbf{l}$$

$$\therefore \qquad E = -\frac{\partial V}{\partial l}$$

Or considering the component in the x-direction, say, we have

$$\delta V = -E_x \, \delta x$$

$$\therefore \qquad E_x = -\frac{\partial V}{\partial x}$$

Similarly,

$$E_y = -\frac{\partial V}{\partial y} \qquad \text{and} \qquad E_z = -\frac{\partial V}{\partial z}$$

$\therefore$ At any point,

$$\begin{aligned} \mathbf{E} &= \mathbf{i}_x E_x + \mathbf{i}_y E_y + \mathbf{i}_z E_z \\ &= -\mathbf{i}_x \left(\frac{\partial V}{\partial x}\right) - \mathbf{i}_y \left(\frac{\partial V}{\partial y}\right) - \mathbf{i}_z \left(\frac{\partial V}{\partial z}\right) \\ &= -\left\{\mathbf{i}_x \left(\frac{\partial V}{\partial x}\right) + \mathbf{i}_y \left(\frac{\partial V}{\partial y}\right) + \mathbf{i}_z \left(\frac{\partial V}{\partial z}\right)\right\} \\ &= -\text{grad } V = -\nabla V \end{aligned} \tag{1.30}$$

That is, **E** is the maximum rate of change of V at that point. Or alternatively, consider the inverse problem, i.e. to determine the increment of the function V from some point A to another point B. Hence consider any line (Figure 1.10) joining AB, and let l be the distance measured along it from A.

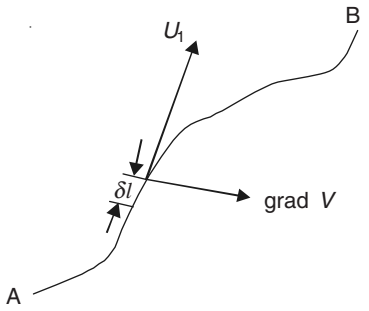


Figure 1.10 Potential difference.

$$\therefore \qquad V_B - V_A = \int_A^B \left(\frac{\partial V}{\partial l}\right) dl$$

But

$$\left(\frac{\partial V}{\partial l}\right) dl = dV = (\text{grad } V) \cdot d\mathbf{l}$$

Hence,

$$V_B - V_A = \int_A^B (\text{grad } V) \cdot d\mathbf{l}$$

that is, the function V is independent of the path traversed. Therefore, over a closed path,

$$\oint_{ABA} (\text{grad}\, V) \cdot d\mathbf{l} = \int_A^B + \int_B^A = (V_B - V_A) + (V_A - V_B)$$

$$= 0$$

Thus it is seen that the potential field is also irrotational. In fact the terms 'conservative' and 'irrotational' are synonymous. Then the question arises as to why both terms are in use when they mean the same thing. Conventional usage has been that irrotational refers normally to velocity fields whilst conservative refers to force fields such as **E** or gravity.

1.6.2 Potential Function and Flux Function

As a help to analyzing problems, 'a potential function V (or ψ)' is defined so that the change in this function between any two points is proportional to the change in the potential between them. Its value at any point, with respect to some origin of potential, is a direct measure of the potential at that point; and a line joining all points having the same value of potential is called an

equipotential line. No work is done in moving a charge along such a line. Since no work is done when a charge is moved in a direction perpendicular to the field strength, the equipotential lines are perpendicular to the flux lines.

Similarly, a flux line is defined such that U (or ϕ) = constant, defines the flux line. Two lines $U = U_0$ and $U = U_0 + n$ have n units of flux passing between them.

These definitions imply that the differences in flux and potential functions represent quantities of flux and potential differences. The zeros of both these functions are chosen arbitrarily. When the lines drawn for constant values of potential and flux functions are so chosen that U and V change in equal steps, then in such a field map the regions enclosed between the successive intersecting lines are 'curvilinear squares'. Since the potential and the flux functions are orthogonal, one function can be derived from the other by using the following equation:

$$\left(\frac{dy}{dx}\right)_{V=\text{constant}} \times \left(\frac{dy}{dx}\right)_{U=\text{constant}} = -1$$

It will be seen later that the flux and the potential functions for the field of a line charge and the field of two charged concentric circular conductors are the same, both fields having concentric circular equipotential lines and straight radial flux lines, i.e.

$$V = \left(\frac{Q}{2\pi\varepsilon_0}\right) \ln r$$

and (1.30 a)

$$U = \left(\frac{Q}{2\pi}\right)\theta$$

where Q is the charge per unit length of the line charge, or is the charge per unit length of the concentric conductors.

1.6.3 Potential Field Expressed as Poisson and Laplace Equations

From Gauss' theorem, we have seen that

$$\text{div } \mathbf{D} = \nabla \cdot \mathbf{D} = \rho_C \text{ (charge density)},$$

and

$$\mathbf{D} = \varepsilon_0 \mathbf{E} \text{ (in free space)}$$

Also, we now have

$$\mathbf{E} = -\text{ grad } V = -(\nabla V)$$

$$\therefore \quad \text{div } \mathbf{E} = \nabla \cdot \mathbf{E} = -\frac{\rho_C}{\varepsilon_0}$$

and

$$\text{div (grad } V) = \nabla \cdot (\nabla V) = -\frac{\rho_C}{\varepsilon_0}$$

or

$$\nabla^2 V = -\frac{\rho_C}{\varepsilon_0} \tag{1.31}$$

Equation (1.31) is the Poisson's equation.

If the charge density $\rho_C = 0$, then

$$\nabla^2 V = 0 \tag{1.32}$$

which is the Laplace's equation.

In the rectangular Cartesian coordinate system,

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

Equations (1.31) and (1.32) are the two very useful equations in physics and engineering. In fact, any electrostatic problem is essentially the problem of solving these equations for the given boundary conditions.

1.7 SOME USEFUL EXAMPLES OF CALCULATION OF FIELDS BY GAUSS' THEOREM AND POTENTIALS

1.7.1 A Group of Charged Particles

From Eq. (1.24) and the corresponding Figure 1.8, the work done due to a single charged particle, in moving a unit charge from the point A to the point B is given by

$$V_B - V_A = \frac{Q}{4\pi\varepsilon_0}\left(\frac{1}{r_B} - \frac{1}{r_A}\right)$$

If there are n charges in the system, then the work done is (Figure 1.11).

$$V_B - V_A = \sum_1^n \left(\frac{Q}{4\pi\varepsilon_0}\right)\left(\frac{1}{r_B} - \frac{1}{r_A}\right) \tag{1.33}$$

As usual we assume the zero potential as being at a very great distance from this group, and

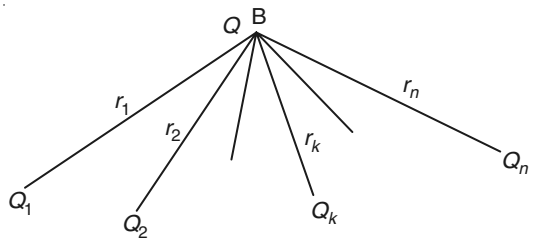


Figure 1.11 A group of charged particles.

allow the point A to move to infinity, i.e. $V_A = 0$, then

$$V_B = \sum_1^n \left(\frac{Q}{4\pi\varepsilon_0}\right)\left(\frac{1}{r_B}\right)$$

or, in general

$$V = \left(\frac{1}{4\pi\varepsilon_0}\right)\sum_1^n \left(\frac{Q}{r}\right) \tag{1.34}$$

1.7.2 A Hollow Charged Sphere of Radius a, and Carrying a Charge Q

In general **E** and **D** can be calculated by the vector addition of contributions from each charge in the field. But when the charges are distributed on the surface of solid bodies, the process of vector addition resolves into that of integration which may not always be an easy one. However, in some cases this can be bypassed by using the Gauss' theorem.

For the charged sphere, we have to find the field at a point P whose distance from the centre of the sphere is r. We think of a Gaussian surface which in this case is a concentric sphere of radius r on which the point P lies (Figure 1.12). This is Σ.

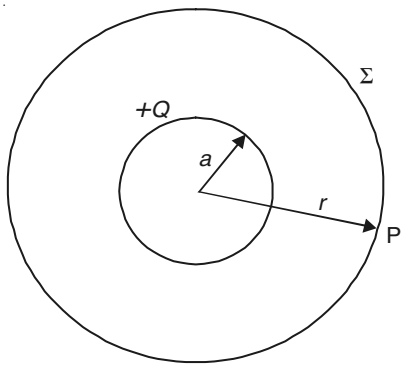


Figure 1.12 A charged sphere.

By Gauss' theorem,

$$\text{Flux of } \mathbf{D} \text{ out of } \Sigma = \text{enclosed charge} = Q$$

Since the spheres are concentric,

$$D \cdot 4\pi r^2 = Q$$

$$\therefore \qquad \mathbf{D} = \left(\frac{Q}{4\pi r^2}\right)\mathbf{r}_1 \qquad \text{and} \qquad \mathbf{E} = \left(\frac{Q}{4\pi\varepsilon_0 r^2}\right)\mathbf{r}_1 \tag{1.35}$$

This is for points outside the sphere (i.e. $r \geq a$). For points inside the hollow sphere, applying a Gaussian surface inside the sphere, charge enclosed = 0.

$$\therefore \qquad E_i = 0 \qquad \text{and} \qquad D_i = 0 \tag{1.36}$$

It is to be noted that for external points the sphere behaves as if all the charges are located at the centre of the sphere.

The next step is to find the potential and from Eqs. (1.35) and (1.36), it is evident that E is a function of r only.

$$\therefore \qquad V = -\int \mathbf{E} \cdot d\mathbf{l}$$

$$= -\int_{\infty}^{r} \frac{Q}{4\pi\varepsilon_0 r^2}\, dr, \text{ for any point external to the sphere.}$$

$$= -\frac{Q}{4\pi\varepsilon_0}\int_{\infty}^{r}\frac{dr}{r^2} = -\frac{Q}{4\pi\varepsilon_0}\left[\frac{-1}{r}\right]_{\infty}^{r} = \frac{Q}{4\pi\varepsilon_0 r} \tag{1.37}$$

For internal points, since $E_i = 0$, only the constant of integration remains, which from the interface condition at $r = a$, is

$$V_a = \frac{Q}{4\pi\varepsilon_0 a} = V_i \tag{1.38}$$

that is, all the points inside the sphere are at the same potential which is equal to that on the surface.

It is also possible to evaluate the potential distribution, ab initio, without determining the complete **E** field. Consider an elemental strip of charge distribution on the surface of the sphere, as shown in Figure 1.13.

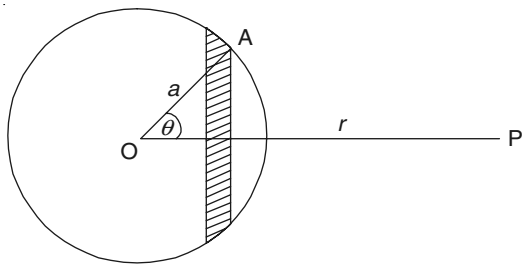


Figure 1.13 Elemental charge distribution on the sphere.

Charge on the hatched strip is

$$\frac{Q}{4\pi a^2}(2\pi a \sin\theta)(a\,\delta\theta) = \frac{Q}{2}\sin\theta\,\delta\theta$$

and the length

$$\mathrm{AP} = (r^2 + a^2 - 2ra\cos\theta)^{1/2}$$

Hence the contribution of the strip to the potential at P is

$$\frac{\left(\frac{Q}{2}\right)\sin\theta\,\delta\theta}{4\pi\varepsilon_0 (r^2 + a^2 - 2ra\cos\theta)^{1/2}} \tag{1.39}$$

Therefore, adding up all such contributions,

$$V_{\mathrm{P}} = \frac{Q}{8\pi\varepsilon_0}\int_0^{\pi}\left(\frac{\sin\theta\,d\theta}{(r^2 + a^2 - ra\cos\theta)^{1/2}}\right) \tag{1.40}$$

On integrating, the answers come out the same as in Eqs. (1.37) and (1.38). This problem shows how the choice of a suitable Gaussian surface simplifies the method of solving it.

1.7.3 Uniformly Distributed Charge on an Infinite Circular Cylinder

Let Q be the charge on the surface of the cylinder of radius a, per unit length. The suitable Gaussian surface would be a concentric cylinder of radius r and of unit length (Figure 1.14). This is Σ. (The end effects have been neglected.) By symmetry, the lines of force are radial, and **E** and **D** are functions of r only.

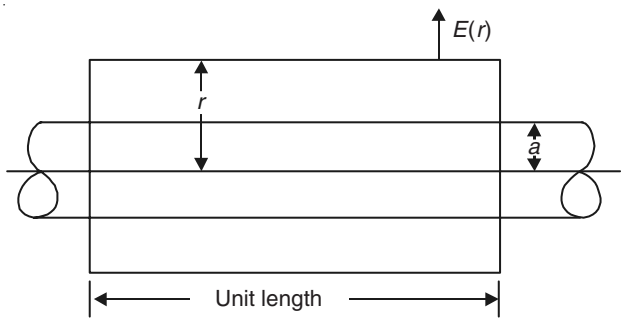


Figure 1.14 Infinitely long charged cylinder.

By Gauss' theorem,

$$\text{Flux of } \mathbf{D} \text{ out of } \Sigma = \text{enclosed charge} = Q \text{ coulombs/metre}$$

Since the cylinders are coaxial,

$$D \cdot 2\pi r \cdot 1 = Q, \qquad l = 1$$

$$\therefore \qquad \mathbf{D} = \left(\frac{Q}{2\pi r}\right)\mathbf{r}_1 \qquad \text{and} \qquad \mathbf{E} = \left(\frac{Q}{2\pi\varepsilon_0 r}\right)\mathbf{r}_1 \tag{1.41}$$

for points outside the cylinder ($r \geq a$).

For points inside the hollow cylinder, applying a similar Gaussian surface, the charge enclosed = 0.

$$\therefore \qquad \mathbf{D}_i = 0 \qquad \text{and} \qquad \mathbf{E}_i = 0 \tag{1.42}$$

The next step is to evaluate the potential,

$$V = -\int \mathbf{E}\cdot d\mathbf{l} = -\int_{r}^{r_R} \frac{Q}{2\pi\varepsilon_0 r}\, dr = -\left(\frac{Q}{2\pi\varepsilon_0}\right)\ln\left(\frac{r_R}{r}\right) \tag{1.43}$$

where r_R is a reference point at a distance r_R from the axis.

Note that the limit of the integral cannot be taken as ∞, as in this problem the cylinder is taken to be infinitely long and hence the charges exist at ∞.

On the surface of the cylinder,

$$V_a = \left(-\frac{Q}{2\pi\varepsilon_0}\right)\ln\left(\frac{r_R}{a}\right) \tag{1.44}$$

On the inside of the cylinder $\mathbf{E}_i = 0$, and hence for all inside points, $V = V_a$.

1.7.4 Infinitely Long Straight Line Charge

While a point charge may be realized in an electron, an infinitely long line charge would be a figment of imagination. Yet the practical importance is that it represents the problem of the charged wire in overhead transmission lines. The line carries a charge of Q C/m. The suitable Gaussian surface is a concentric cylinder of radius r and of unit axial length (Figure 1.15). The end effects have been neglected.

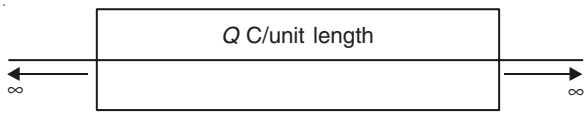


Figure 1.15 Infinitely long line charge.

By symmetry, **D** is radial and normal to the axis of the cylinder.

$$\therefore \qquad \mathrm{D} \cdot 2\pi r \cdot 1 = Q$$

$$\therefore \qquad \mathbf{D} = \left(\frac{Q}{2\pi r}\right)\mathbf{r}_1 \quad \text{and} \quad \mathbf{E} = \left(\frac{Q}{2\pi\varepsilon_0 r}\right)\mathbf{r}_1 \tag{1.45}$$

To evaluate the potential,

$$V = -\int_{\infty}^{r} \mathbf{E} \cdot d\mathbf{r} = -\left(\frac{Q}{2\pi\varepsilon_0}\right)\ln r \tag{1.46}$$

It should be *noted* that the above expression is that *part* of V which varies with r. There will be an infinite constant of integration. Again the reason for this is that the line charge extends to infinity and so the absolute potential cannot be defined as the work done in bringing a unit charge from infinity. Equipotential surfaces will be coaxial cylinders and on a normal plane these would intersect as concentric circles.

1.7.5 A Group of Parallel Line Charges

By using the principle of superposition (Figure 1.16), we obtain

$$V_{\mathrm{P}} = -\left(\frac{1}{2\pi\varepsilon_0}\right)\left(Q_1 \ln r_1 + Q_2 \ln r_2 + Q_3 \ln r_3 + \cdots + Q_n \ln r_n\right) \tag{1.47}$$

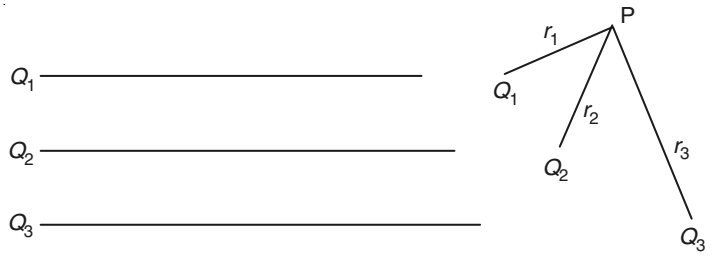


Figure 1.16 A group of parallel line charges.

Consider the special case of when there are two line charges,

$$V_P = -\left(\frac{1}{2\pi\varepsilon_0}\right)\left(Q_1 \ln r_1 + Q_2 \ln r_2\right) \tag{1.48}$$

A particular case of this is when the two charges are equal and opposite, i.e. $Q_1 = Q$ and $Q_2 = -Q$. Then,

$$V_P = -\left(\frac{Q}{2\pi\varepsilon_0}\right)\left(\ln r_1 - \ln r_2\right)$$

$$= \left(\frac{Q}{2\pi\varepsilon_0}\right)\ln\left(\frac{r_2}{r_1}\right) \tag{1.49}$$

The equipotentials will be given by

$$\frac{r_2}{r_1} = \text{constant} \tag{1.50}$$

1.7.6 Charges Distributed Uniformly Over an Infinitely Plane Surface

Consider an infinite plane surface uniformly charged with a surface charge density of ρ_S. By symmetry the lines of force are normal to this surface and **E** is a function of the distance normal to this plane, and $\mathbf{E}(x) = -\mathbf{E}(-x)$. The Gaussian surface is a cylindrical surface bisected by the charged plane as shown in Figure 1.17.

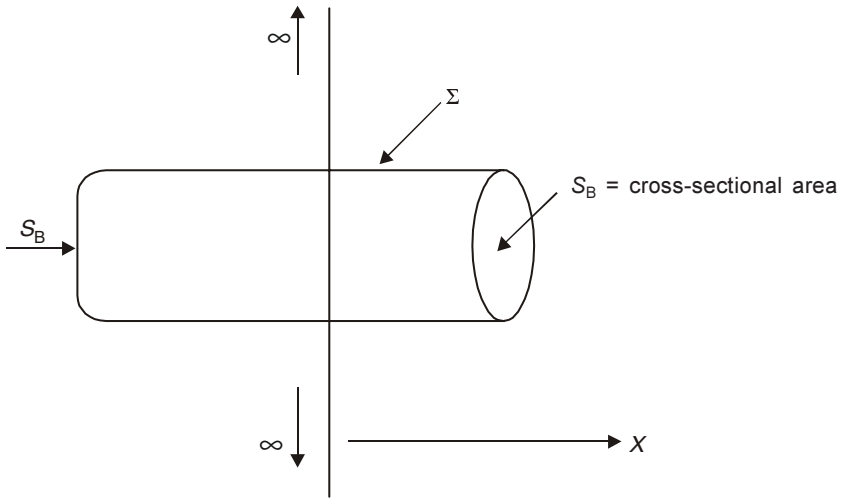


Figure 1.17 Uniformly charged infinite plane.

Thus,

$$\oint_{\Sigma} \mathbf{D} \cdot d\mathbf{S} = 2D(x)S_B = \rho_S S_B$$

$$\therefore \qquad \mathbf{D}(x) = \mathbf{i}_x \left(\frac{\rho_S}{2} \right) \qquad \text{and} \qquad \mathbf{E}(x) = \mathbf{i}_x \left(\frac{\rho_S}{2\varepsilon_0} \right) \tag{1.51}$$

$\therefore$ The electric field is constant and does not depend on the distance of the point from the charged sheet. As in the previous example (Section 1.7.5), it can be shown that the finite part of the potential is

$$V = -\frac{\rho_S x}{2\varepsilon_0} \tag{1.52}$$

Considering a special case of the two parallel planes (Figure 1.18) which are oppositely charged, i.e. $+\rho_S$ and $-\rho_S$, we get $\mathbf{E}$ in the space between the two planes as

$$\mathbf{E} = \frac{\rho_S}{\varepsilon_0} \qquad \text{and} \qquad \mathbf{E}(\text{outside}) = 0 \tag{1.53}$$

Again it should be noted that the infinite charged plane is an ideal concept, and yet the above representation simulates the parallel plate capacitor away from its edges.

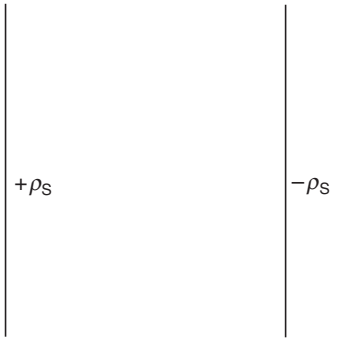


Figure 1.18 Two parallel charged planes.

1.7.7 Electric Dipole

An electric dipole (Figure 1.19) is a pair of equal and opposite charges held apart at a spacing which is small compared with the distances at which the field is calculated.

Let the point charges be $\pm Q$ and the distance between them be $2d$, then

$$m = \text{moment of the dipole} = 2dQ$$

The potential at the point P is

$$V_P = \left(\frac{Q}{4\pi\varepsilon_0} \right) \left(\frac{1}{PQ} - \frac{1}{PQ'} \right)$$

where

$$PQ = r - d\cos\theta \qquad \text{and} \qquad PQ' = r + d\cos\theta$$

$$\therefore \qquad V_P = \left(\frac{Q}{4\pi\varepsilon_0} \right) \left(\frac{1}{r - d\cos\theta} - \frac{1}{r + d\cos\theta} \right)$$

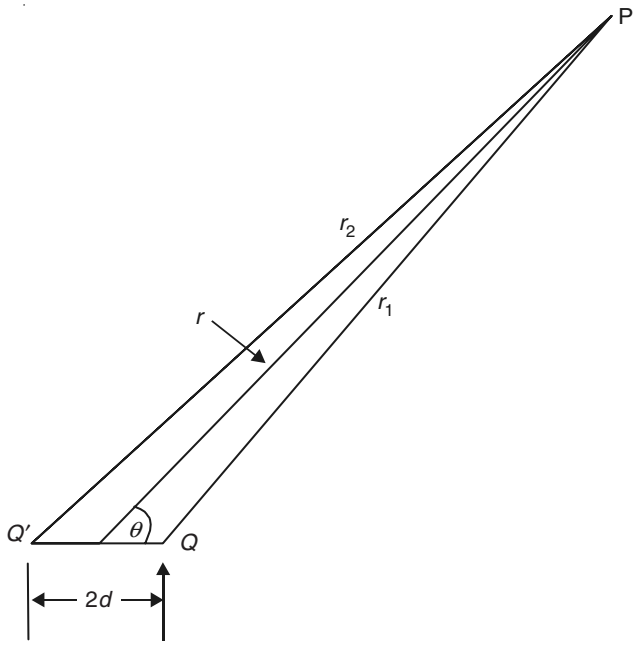


Figure 1.19 Electric dipole ($r >> d$).

$$= \left(\frac{Q}{4\pi\varepsilon_0}\right)\left(\frac{2d\cos\theta}{r^2 - d^2\cos^2\theta}\right)$$

In the limit,

$$V_P = \frac{m\cos\theta}{4\pi\varepsilon_0 r^2} \tag{1.54}$$

To evaluate **E**, we use the spherical coordinate system, with its origin at the centre of the dipole. Hence,

$$E_r = -\frac{\partial V}{\partial r} = +\frac{m\cos\theta}{2\pi\varepsilon_0 r^3}$$

$$E_\theta = -\left(\frac{1}{r}\right)\left(\frac{\partial V}{\partial\theta}\right) = +\frac{m\sin\theta}{4\pi\varepsilon_0 r^3} \tag{1.55}$$

Thus, the **E** field falls as the inverse cube of the distance.

The chief engineering application of the dipole is in the theory of aerials. It also has other applications in physics.

PROBLEMS

1.1 What maximum charge can be put on a sphere of radius 1 m, if the breakdown of air is to be avoided.

(For breakdown of air, $|\mathbf{E}| = 3 \times 10^6$ V/m)

Ans.: $Q_{max} = 3.3 \times 10^{-4}$ C

1.2 Eight identical spherical drops of mercury charged to 12 V above the earth potential are made to coalasce into a single spherical drop. What is the new potential?

Ans.: $V_2 = 48$ V

1.3 An infinite plane sheet of charge gives an electric field $\sigma/(2\varepsilon_0)$ at a point P, a distance a from it. Show that half the field is contributed by charge whose distance from P is less than $2a$; and that in general all but f per cent of the field is contributed by charge whose distance from P is less than $100a/f$.

1.4 Two point charges Q_1 and Q_2 are located at the origin and the point $(x_2, y_2, 0)$, respectively. Find the force on the charge Q_1, expressed in newtons, if Q_1 and Q_2 are in μC, and the distances is in metres.

1.5 Two concentric circular cylinders of radii r_1 m and r_2 m, respectively, carry uniform surface charges of ρ_{S1} and ρ_{S2} μC respectively. Both the cylinders carrying the charges extend to infinity along the z-axis ($\pm \infty$). Specify the necessary Gaussian surfaces required to evaluate the electric flux density in all the regions.

1.6 Derive Coulomb's law, starting from Gauss' theorem. State any reasonable assumptions which you think are necessary for the derivation.

2 Conductors and Insulators in Electrostatic Field

2.1 CONDUCTORS AND INSULATORS

So far we have discussed the interactions between charged particles or bodies without referring to the properties of the bodies themselves. From the point of view of electrostatics we can classify all substances according to the ease with which electric charges move in them. Thus charges can be transferred from point to point in metals, but not in dry cotton or rubber. Thus the metal is a conductor of electricity whilst the other type is described as an insulator. The difference between the two is that, in metals or conductors, the electrons are easily detached from the atoms; in the insulators, the detachment is difficult. To illustrate the point, if a piece of metal wire is placed lengthwise in field, a potential difference would be set up between the ends of the wire. This would cause the free electrons to move along the length of the wire. There would be interruptions due to collisions with atoms, but a general drift of electrons sets in and an electric current starts flowing in the wire. This happens extremely quickly, possibly in a millionth of a second. The wire thus behaves like a pipe along which charge can flow from one end to the other. The terms conductor and insulator are relative, and would depend on the type and the duration of the experiments. However metals, carbon, aqueous salt solutions, etc. can be considered as good conductors and dry gases as good insulators.

2.2 CONDUCTORS IN THE ELECTROSTATIC FIELD

Since the charges in a conductor are mobile, even a small electric field applied to any conductor would set the charges in motion. Hence if a field is truly electrostatic, then there can be no electric force in or within the conducting body. So for a conductor under electrostatic conditions, there can be no potential difference in it. Thus the conductor surface is an equipotential surface and the whole interior would be at the same potential.

At this point we remind ourselves about 'the constant of a medium'. We have already stated that for vacuo and air, this is the 'absolute permittivity of free space' denoted by ε_0. For all media, in general, this is expressed as

$$\mathbf{D} = \varepsilon \mathbf{E} \tag{2.1}$$

where

$$\varepsilon = \varepsilon_0 \varepsilon_r \tag{2.2}$$

ε_r = the relative permittivity of the medium.

We shall discuss about ε_r in detail when we consider the properties of insulators. In conductors, $\varepsilon_r = 1$.

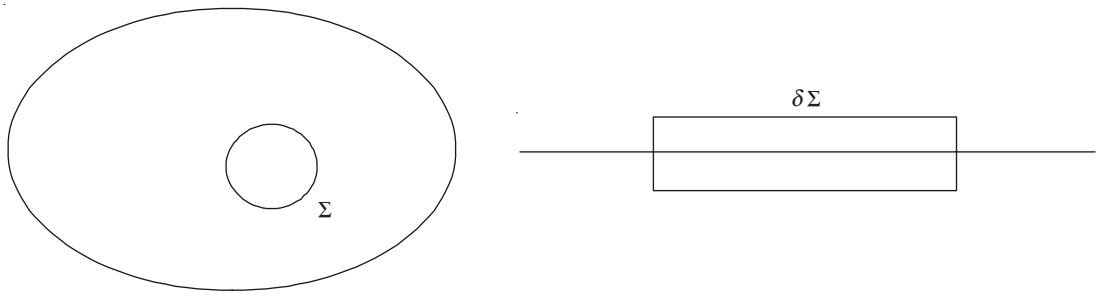
Going back to the conducting body under consideration, let us consider a Gaussian surface Σ inside it [Figure 2.1(a)]. Because $\mathbf{E} = 0$ inside the conductor, the flux of $\mathbf{E}$ out of $\Sigma = 0$. Hence by Gauss' theorem,

$$\nabla \cdot \mathbf{E} = \frac{Q}{\varepsilon} \tag{2.3}$$

where Q is the charge enclosed by Σ. Therefore, either ε is infinite or Q is zero. Infinite ε would mean that each charge is surrounded by conductor so that $\mathbf{E} = 0$ (i.e. no electric force). Zero Q means that there is no charge within the body of the conductor but there can be charges on the conductor surface. This has been verified experimentally. Also, indirect evidence shows that ε is finite. Therefore, inside the conducting material

$$\mathbf{E} = 0 \qquad \text{and} \qquad \mathbf{D} = 0 \tag{2.4}$$

Now let us consider a coin-shaped Gaussian surface on the surface of the conductor [Figure 2.1(b)]



(a) Conductor body (b) Conductor surface

Figure 2.1 Gaussian surfaces inside and on the surface of a conducting body.

such that one of its flat surfaces parallel to the conductor surface lies just inside the conductor and the other face just outside it. If the area of its flat surface is $\delta\Sigma$ and the surface charge density is σ, then the charge enclosed by this surface element is ($\sigma\,\delta\Sigma$). For the Gaussian surface under consideration, $\mathbf{D}$ is zero on the inner flat surface, no contributions across the edge, and the only contribution is from the outer flat surface.

$$\therefore \qquad D = \sigma \qquad \text{and} \qquad E = \frac{\sigma}{\varepsilon} \tag{2.5}$$

Thus charges on conductors lie entirely on their surfaces and distribute themselves so as to make the conductor surfaces equipotential. Electric flux emerges normally from the positively charged surfaces and enters the negatively charged surfaces similarly.

Next, by considering a tubular Gaussian surface whose sides are lines of force and whose ends are within the two conductors, we find that there is no flux of $\mathbf{D}$ from such a surface, because

D is parallel to the sides and is zero at the ends. Hence this surface encloses no charge. Such a tube which combines an element of positive charge to an equal element of negative charge is called a 'Faraday tube'. There is a continuous fall in potential as one traverses this tube from the positive end to its negative end.

2.3 RELATION BETWEEN ELECTROSTATIC POTENTIAL AND CHARGES ON CONDUCTING BODIES

2.3.1 The Case of an Isolated Conductor

Let us consider an isolated charged body. The charges would be so distributed on the body that its surface would be equipotential, and inside it **E** = 0. If the total charges on the body are increased, the redistribution of the charges would be such that the above conditions still get fulfilled. Thus, *the relative distribution of charges on an isolated conducting body is independent of the total charge on the body*. The potential of the isolated conducting body would depend on the total charge on this body. Thus if Q is the total charge on the body and V its corresponding potential, then

$$Q = CV \tag{2.6}$$

where C is a constant depending on the shape of the body and is called the *capacitance* of the body, characterizing the 'capacity' of the isolated body with respect to the charge accumulated per unit of potential.

Equation (2.6) defines the capacitance of an isolated conductor as well as shows a method of determining the capacitance. As a simple example, we know that the potential of a spherical conductor with charge Q on it is given by [see Eq. (1.38)]

$$V = \frac{Q}{4\pi\varepsilon_0 r}$$

where r is the radius of the sphere.

$$\therefore \qquad \text{Its capacitance} = C = \frac{Q}{V} = 4\pi\varepsilon_0 r \tag{2.7}$$

Capacitance is a very important concept and its unit is coulomb/volt (C/V) and is called farad (F).

2.3.2 The Case of Two Bodies with Equal Charges of Opposite Signs

This is a system which is often encountered in practice. Charges of equal magnitude but of opposite signs on two bodies would be obtained if the two bodies are connected to a source of electromotive force which pumps charges of one sign from one body to the other. Thus if a charge of $+Q$ is taken from one body and transferred to another body, then a charge of $-Q$ is left on the first body. In practice such a system is called 'capacitor' and the two conducting bodies of the system are referred to as the 'electrodes'. A very simple relationship between the charge Q and the potential difference between its electrodes (= V) exists. This is written as

$$Q = CV = C(V_1 - V_2) \tag{2.8}$$

where V_1 and V_2 are the potentials of the two electrodes, respectively. As mentioned in Section 2.3.1, the unit of capacitance C is farad (= coulomb/volt).

2.3.3 Methods of Evaluating Capacitance (C)

Equation (2.8) tells us how to evaluate C in addition to giving its definition. There are basically two methods of evaluating C. The steps are as follows:

METHOD A

Assume charges $\pm Q$ on the two electrodes of the capacitor.

By using the Gauss' theorem, find **D** and hence **E**.

Find the potential difference (P.D.) between the two electrodes, which is given by $\int \mathbf{E} \cdot d\mathbf{l}$.

Hence,

$$C = \frac{Q}{\text{P.D.}}$$

METHOD B

Assume a potential difference (P.D.) between the electrodes and hence the potential distribution.

Find **E** from the equation $\mathbf{E} = -\text{ grad } V$, and hence **D**.

By using the Gauss' theorem, find Q—the enclosed charge on one of the electrodes.

Hence,

$$C = \frac{\text{charge}}{\text{P.D.}}$$

The next logical step would now be to discuss the capacitance of a number of different arrangements of electrodes. But before we do that, a further generalization is possible, i.e. so far we have considered the intervening space between the two electrodes to be either vacuum or air which have the permittivity as ε_0. However, it is possible to have this space or gap filled with any insulating material. The effect of this change would be to change the 'free space constant ε_0' by the permittivity of the new material. Before we do make this change, we shall now study the behaviour of insulators under the effect of electric fields and then come back to the calculation of capacitances.

2.4 THE BEHAVIOUR OF INSULATORS (OR DIELECTRICS) IN A STATIC ELECTRIC FIELD

It has been stated earlier that insulators (or as they are also called dielectrics) are those substances which do not possess free electric charges. This does mean that when they are introduced in an electric field, the field would not get affected by them.

In order to understand the mechanism by which a dielectric material affects an electric field in which it is introduced, we shall look at the behaviour of a single atom of this material. In the absence of the electric field such an atom is electrically neutral. This is strictly true for far points, i.e. for distances which are large compared with the sub-atomic distances. For such small distances, the atom does exert an electric field. So when we look at the atom from large distances, the atom appears to be as if it were made up of two concentric uniformly charged spheres carrying equal charges of opposite signs. Such a system produces zero **E** field at all points outside the

atom [Figure 2.2(a)]. Next we consider in detail what happens when an external electric field **E** falls on this dielectric material. Since the atom of the material consists of positive and negative charges, so, when the field is applied, the positive charges will be pulled in one direction and the negative charges in the opposite direction. In the neutral condition (i.e. when there was no external electric force) the elastic forces in the atom tended to keep the two imaginary charged spheres concentric. When the external field is applied, these two charged spheres tend to move in opposite directions and there is a displacement of their centres which depends on the intensity of the external applied field. Since the centres of the two spheres are no longer at the same point [Figure 2.2(b)], the electric field at distant points is no longer zero. It is now equal to the field

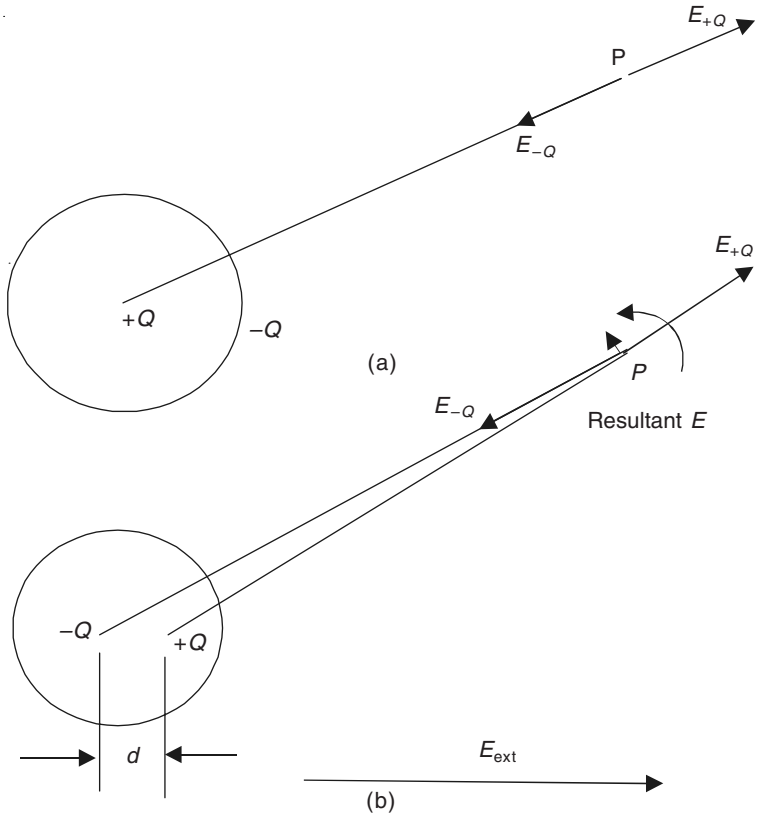


Figure 2.2 (a) Model of a neutral atom; (b) the same atom in an external electric field.

produced by the two point charges at a distance d from one another. It should be noted that this distance d is much smaller than the distance of the observer from the atom. Such a system of two closely-placed point charges of equal magnitude but of opposite signs is called an 'electric dipole' (see Section 1.6.6). Thus the secondary electric field created by a neutral atom of a dielectric in an external electric field is the same as that due to a dipole. So when a dielectric material is introduced in an external electric field, the secondary electric field due to the deformed atoms of the dielectric is the same as the sum of the corresponding electric dipoles situated in a vacuum. In this case the substance is said to be polarized and the process is known as *polarization.*

This type of polarization takes place with a very wide class of dielectrics. But there is

another class of dielectrics in which the molecules are dipoles even when there is no externally applied electric field. Such permanent dipole molecules are called *polar molecules*, an example of which is water. In the absence of any externally applied electric field, these dipoles are so distributed statistically in space that there is no net macroscopic field under normal conditions. When an external electric field is applied, there is a tendency for all dipoles to align themselves in the direction of the applied field, thus creating a secondary electric field which is the same as the electric field due to a large number of electric dipoles. Hence all dielectric substances when brought into an electric field can be considered equivalent to a large number of tiny electric dipoles. The secondary macroscopic electric field due to a polarized dielectric can thus be regarded as the field of these dipoles located in vacuum. Thus the analysis of the behaviour of dielectric substances in an electrostatic field reduces to the analysis of a system of dipoles located in vacuum. So now we have to determine from the potential and the electric field due to a single dipole, to those due to a large aggregate of elemental electric dipoles.

2.4.1 The Potential and Electric Field due to an Aggregate of Dipoles (Polarization Vector)

We have, from Section 1.6.6, the potential at a point, due to a dipole (Q, d) in vacuo as

$$V = \frac{Q\,d\cos\theta}{4\pi\varepsilon_0 r^2}$$

And $\mathbf{p} = Q\mathbf{d}$ is defined as the dipole moment.

If $\mathbf{u}$ is the unit vector in the direction of the point under consideration from the dipole (Figure 2.3), then,

$$V = \frac{\mathbf{p}\cdot\mathbf{u}}{4\pi\varepsilon_0 r^2} \tag{2.9}$$

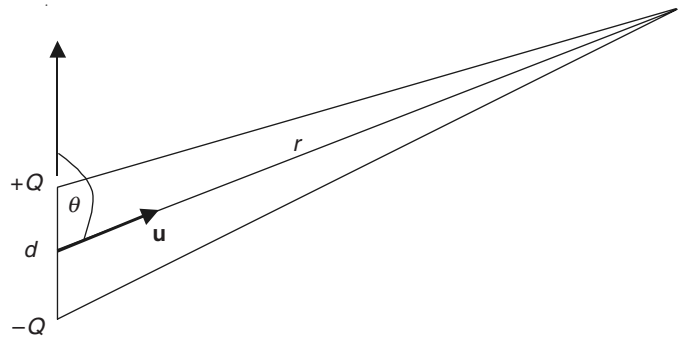


Figure 2.3 An electric dipole.

In a polarized dielectric medium, the density of the electric dipole moment $\mathbf{P}$ is

$$\mathbf{P} = \frac{\sum_{dv} \mathbf{p}}{dv} \tag{2.10}$$

at the volume element dv, due to all polarized atoms in that element.

This is the definition of the 'Polarization Vector'. The electric field intensity due to the dipole is given by [from Eq. (2.9)],

$$E_r = -(\text{grad } V)_r = -\frac{\partial V}{\partial r} = \frac{p \cos\theta}{2\pi\varepsilon_0 r^3} \tag{2.11}$$

$$E_\theta = -(\text{grad } V)_\theta = -\frac{1}{r}\frac{\partial V}{\partial \theta} = \frac{p \sin\theta}{4\pi\varepsilon_0 r^3} \tag{2.12}$$

The dipole moment of the elemental volume dv is

$$d\mathbf{p} = \mathbf{P}\, dv \tag{2.13}$$

$\therefore$ The potential due to the polarization of this element is given by

$$dV = \frac{\mathbf{P} \cdot \mathbf{u}\, dv}{4\pi\varepsilon_0 r^2} \tag{2.14}$$

$\therefore$ The potential due to all the elemental dipoles inside a volume v is

$$V = \frac{1}{4\pi\varepsilon_0} \iiint \frac{(\mathbf{P} \cdot \mathbf{u})\, dv}{r^2} \tag{2.15}$$

Thus **P** at a point is some function of the total **E** field at that point. So we find that the **E** field is a function of the **P** vector at each point and **P** at a point is function of the total **E** field. Then the question is how is **P** to be determined? Experimentally it is found that

$$\mathbf{P} = \varepsilon_0 \chi_e \mathbf{E} \tag{2.16}$$

where χ_e is the electric susceptibility of the medium. It is a characteristic of each dielectric. For non-polar dielectrics, the mechanism of polarization is of internal atomic nature and hence independent of temperature. For polar insulating materials the thermal vibrations of the polarized molecules prevent their alignment with the electric field and hence their susceptibility decreases with temperature. In some crystalline structures, such as quartz, strong internal forces interfere with the forces produced by the externally imposed electric field. The result is that the direction of **P** is not necessarily the same as that of **E**. Such dielectrics are called *anisotropic dielectrics.* However, we shall deal mostly with isotropic materials.

2.4.2 Charge Distribution Equivalent to a Polarized Dielectric

So far we have shown that if viewed macroscopically, a polarized substance can be considered as an aggregate of elemental electric charges located in vacuum. It can also be shown that the secondary electric field of these elemental electric dipoles can be ascribed to a certain volume and surface distributions of charges situated in vacuum. For this purpose, consider a surface element $d\mathbf{S}$ in a polarized dielectric (Figure 2.4). Let the polarization vector in the neighbourhood of $d\mathbf{S}$ be **P,** the angle between the direction of the two being α. In the process of polarization, this surface $d\mathbf{S}$ is crossed by a number of positive charges in one direction, and a certain number of negative charges in the opposite direction. Let the displacement of the positive charges from their equilibrium point be d_+ and that of the negative charges in the opposite direction be d_-. Then

$$d_+ + d_- = d$$

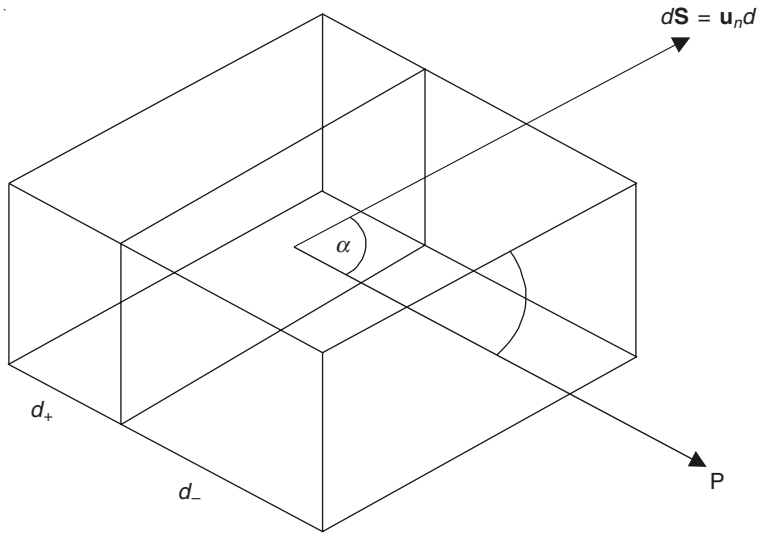


Figure 2.4 Movement of polarization charges for a distance ($d_+ + d_-$)

If N is the number of molecules per unit volume, then the total charge that crossed the surface $d\mathbf{S}$ in its positive direction in the process of polarization is

$$\begin{aligned} dQ &= N(+Q)\, d_+ d\mathbf{S} \cos\alpha - N(-Q) d_- d\mathbf{S} \cos\alpha \\ &= NQd\, d\mathbf{S} \cos\alpha = NQ\mathbf{d} \cdot d\mathbf{S} \end{aligned} \tag{2.17}$$

But by definition, $\mathbf{P} = NQ\mathbf{d}$, and

$$\therefore \quad dQ = \mathbf{P} \cdot d\mathbf{S} \tag{2.18}$$

Now, let us consider a very small closed surface ΔS, then the total positive charge that left the volume enclosed by ΔS $(= \Delta v)$ in the process of polarization is

$$\Delta Q = \oiint_{\Delta S} \mathbf{P} \cdot d\mathbf{S} \tag{2.19}$$

The charge that remained inside ΔS is $-\Delta Q$. The density of this charge is

$$\rho_P = -\frac{\Delta Q}{\Delta v} = -\frac{\oiint_{\Delta S} \mathbf{P} \cdot d\mathbf{S}}{\Delta v} \tag{2.20}$$

In the limit the above expression becomes,

$$\rho_P = -\operatorname{div} \mathbf{P} = -\nabla \cdot \mathbf{P} \tag{2.21}$$

This is the volume density of the 'bound' charges. It should be noted that ρ_P is not the density of the free electric charges, but the mean density of the charges firmly bound to their atoms. In the process of polarization the balance of the positive and the negative charges in a physically small volume gets disturbed. Because the polarization varies from point to point, more charges might leave the volume than enter it. Charges thus created in a polarized dielectric are called 'polarization charges'. These are also referred to as 'bound charges' to denote the fact that they are not freely movable.

Similarly on the boundary surface of a dielectric and vacuum there would exist uncompensated bound charges existing in a thin layer. By analogy with Eq. (2.21), this surface charge density is referred to as the surface density of polarization or of bound charges (= ρ_{SP}). This analysis can similarly be further extended to the boundary surface of two different dielectrics.

2.5 GENERALIZED FORM OF GAUSS' THEOREM

We have thus obtained the electric field intensity created by free and polarization charges according to the same formulae which have been derived for the electrostatic field of charges situated in vacuum. Hence all the deductions discussed in Chapter 1 remain valid here as well. So replacing the charge density ρ by $(\rho + \rho_P)$ and the surface charge density ρ_S by $(\rho_S + \rho_{SP})$, Gauss' theorem in integral form becomes:

$$\oiint_S \mathbf{E} \cdot d\mathbf{S} = \frac{1}{\varepsilon_0} \iiint_v (\rho + \rho_P) dv \tag{2.22}$$

where v is the volume enclosed by the closed surface S.

Since $\rho_P = -$ div **P,** by substituting in the above equaton, we have

$$\oiint_S \varepsilon_0 \mathbf{E} \cdot d\mathbf{S} + \iiint_v \text{div}\mathbf{P} dv = \iiint_v \rho dv \tag{2.23}$$

By using the divergence theorem,

$$\oiint_S (\varepsilon_0 \mathbf{E} + \mathbf{P}) \cdot d\mathbf{S} = \iiint_v \rho dv \tag{2.24}$$

This is the Generalized Gauss' theorem in integral form.

Transforming the surface integral into a volume integral,

$$\iiint_v \text{div}\,(\varepsilon_0 \mathbf{E} + \mathbf{P})\, dv = \iiint_v \rho dv$$

$$\text{div}\,(\varepsilon_0 \mathbf{E} + \mathbf{P}) = \rho \tag{2.25}$$

where ρ is the free charge in the medium.

This is the Gauss' theorem in differential form; and it shows that the vector $(\varepsilon_0 \mathbf{E} + \mathbf{P})$ has the remarkable property that its flux through any closed surface is equal to the total 'free' charges enclosed by that surface. So now we redefine the electric displacement vector **D** as

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P} \tag{2.26}$$

whose validity is completely general. Thus the Gauss' theorem can be written in terms of the general definition of the displacement vector as

$$\oiint_S \mathbf{D} \cdot d\mathbf{S} = \iiint_v \rho\, dv \tag{2.27}$$

and

$$\text{div}\ \mathbf{D} = \rho \tag{2.28}$$

When **P** can be expressed in terms of **E**, then

$$\mathbf{D} = (\varepsilon_0\mathbf{E} + \mathbf{P}) = \varepsilon_0(1 + \chi_e)\mathbf{E} \tag{2.29}$$

The sum $(1 + \chi_e)$ is referred to as the *dielectric constant* or the *relative permittivity* and is denoted by ε_r. Thus,

$$\varepsilon_r = 1 + \chi_e \tag{2.30}$$

The product $\varepsilon_0 \varepsilon_r$ is called the permittivity of the dielectric and its notation is ε, i.e.

$$\varepsilon = \varepsilon_0 \varepsilon_r = \varepsilon_0(1 + \chi_e) \tag{2.31}$$

This is the definition of permittivity. For linear dielectrics,

$$\mathbf{D} = \varepsilon_0 \varepsilon_r \mathbf{E} = \varepsilon\mathbf{E} \tag{2.31a}$$

Some of the relative permittivities of common dielectrics are listed in Table 2.1.

Table 2.1 Relative permittivity of some dielectrics

Medium	*Relative permittivity*
Vacuum	1.0000
Air at 1 atm.	1.00059
Transformer oil	2.24
Distilled water at °C	88.0
Distilled water at 20°C	80.0
Glass	3.8–6.8
Mica	6.0
Rubber	2–3.5
Nylon	3.5

Distilled water is a typical example of a polar dielectric.

2.6 SOME PHYSICAL PROPERTIES OF DIELECTRICS

A brief account of some of the macroscopic properties of dielectrics is given here.

2.6.1 Dielectric Strength

This is the lowest value of the electric field intensity for which a dielectric loses its insulating property and becomes a conductor.

There is a certain critical value of the electric field intensity for every dielectric. For field intensities less than this critical one, the dielectric behaves as an insulator and for intensities greater than the critical value the dielectric behaves as a conductor. This behaviour can be easily understood if we remember that under normal conditions the force between the electrons and the nuclei in an insulator is sufficient to bind them together. When the insulator is introduced in an electrostatic field of sufficient intensity, the electric force acting on some of the electrons may be large enough to tear them away from the rest of the atom. Thus free electrons (or electric charges) appear in the structure of the dielectric resulting in large electric currents. This is known as the

'dielectric breakdown'. Representative values of dielectric strength of some common dielectrics are given in Table 2.2. For gases this value depends on the method of the experiment.

Table 2.2 Dielectric strength of some common dielectrics in kV/mm

Dielectric	*Dielectric strength*	*Dielectric*	*Dielectric strength*
Air	3.20	Steatite	25–40
N_2	3.30	Rubber	40
O_2	2.90	Mica	50
CO_2	2.90	Thermoplasts LDPE	75
H_2	1.90	HDPE	100
He	1.00	XLPE	50
Glass	10–50	Hard PVC	30
Quartz porcelain	20–40	Soft PVC	10
Clay porcelain	20–40		

It also happens that **E** is large enough in certain parts of the dielectric, in which case there may be local sparking or discharging currents (known as *corona*) instead of general dielectric breakdown. However these are undesirable phenomena, though there are some practical useful applications of this type of breakdown. Some examples are lightening rods and provision of sharp edges on aeroplanes.

2.6.2 Dielectric Relaxation

In some substances having polar molecules, the alignment of dipoles under the influence of an external electric field and also the return to their initial statistical distribution after the field is switched off, requires a certain period of time. This time is referred to as the 'relaxation time.' This relaxation time varies greatly for different substances. There are several practical important consequences of this phenomenon. An example is the time required for charging and discharging capacitors with such dielectrics.

2.6.3 Triboelectricity

Historically the first experiment with electric phenomena was performed by rubbing a piece of amber with some type of cloth. The result was that both the pieces of amber and the cloth attracted light pieces of various substances. It is now known that this attraction is due to the appearance of electric charges on the surfaces of the amber and the cloth. This phenomenon is called *triboelectricity* (the word is of Greek origin, meaning electricity created by rubbing), and is not confined only to amber and cloth. Also the rubbing does not produce electricity, but merely promotes it. A likely explanation of this mechanism is that when two different insulators are rubbed, at the surface of the contact the outer electrons in the molecules may be acted upon by additional forces due to nearby molecules of a different type. This force acts against the binding forces in the molecules and tears the electrons away from the original molecule. So, when the two pieces of insulators are separated, they are charged.

2.7 CAPACITANCE: CAPACITORS

Capacitors have already been defined in Section 2.3.2. The capacitor is a system in which two conductors carrying equal and opposite charges are separated by non-conducting space. Initially this space was considered as vacuum or air. Now following our study of dielectrics, we see that the system can be further generalized by allowing the intervening space to be occupied by any insulating material (or dielectric). The method of calculating the capacitance of the system, which is going to be our present topic of discussion, will not be affected by this generalization as only the absolute permittivity ε_0 needs to be replaced by $\varepsilon_0\varepsilon_r$ $(= \varepsilon)$, because the free space in the gap has been replaced by a material substance which would contain bound charges. The effects of these bound charges are adequately accounted for by the above substitution. Figures 2.5 and 2.6

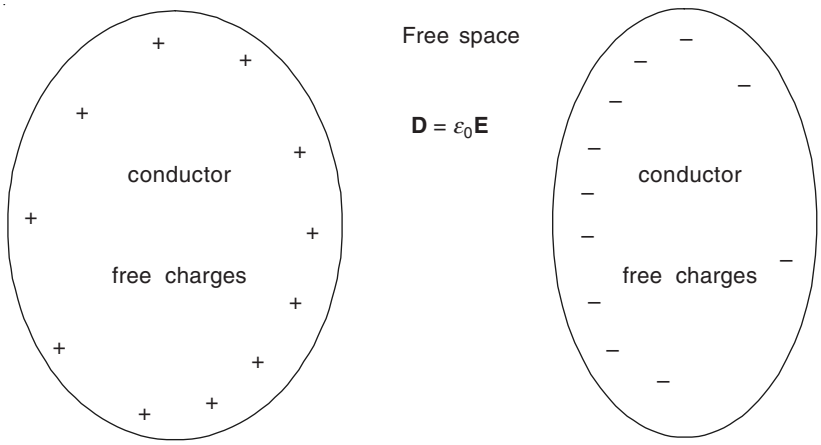


Figure 2.5 Free charges on a two-conductor system.

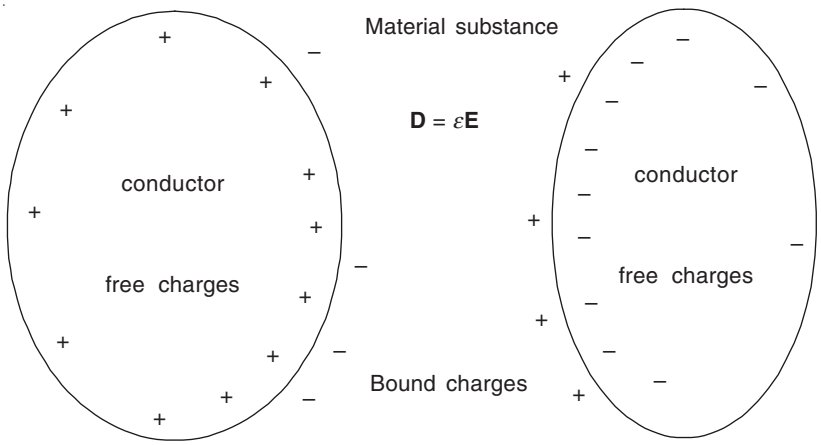


Figure 2.6 Free charges on the conductor surfaces and bound charges on dielectric surfaces.

pictorially explain this modification. It should be noted that at any point on the boundary surface the density of the bound charges would be less than the density of free charges. We shall now discuss a number of different arrangements of capacitors. A point to be noted regarding the

substitution of ε_0 by ε to take account of the effects of the dielectric is really equivalent to accounting for all charges, both free and bound, in the material of the space by a summation or integration effect.

2.8 CALCULATION OF CAPACITANCE

As will be seen later, the capacitors are also devices for storing the electric or the electrostatic energy.

2.8.1 Parallel Plate Capacitor

Consider plates of area A, separation d, with a dielectric of permittivity $\varepsilon = \varepsilon_0\varepsilon_r$. (See Figure 2.7.) Take a small element of surface δS on one plate away from the disturbing effects of the edges.

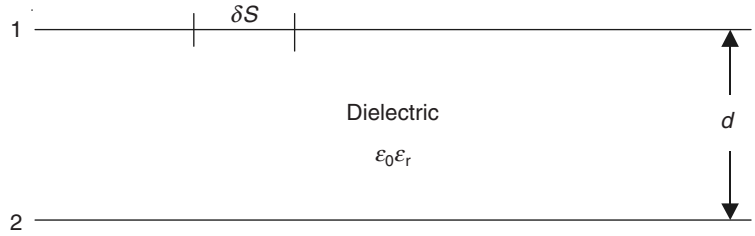


Figure 2.7 A parallel plate capacitor.

The tubes of flux traverse the dielectric uniformly across its thickness from one plate to the other. If the uniform charge density on the plates is denoted by σ, then by Gauss' theorem,

$$D = \sigma \tag{2.32}$$

on the plates as well as across the dielectric.

$$\therefore \quad E = \frac{D}{\varepsilon} = \frac{\sigma}{\varepsilon} \tag{2.33}$$

$\therefore$ The P.D. between the plates = $E \times$ the length of the gap, because E is constant across the gap.

$$\therefore \quad \text{P.D.} = \frac{\sigma d}{\varepsilon} \tag{2.34}$$

$$\therefore \quad \text{Capacitance, } C = \frac{\text{charge}}{\text{P.D.}} = \frac{\sigma A}{\sigma d/\varepsilon} = \frac{\varepsilon A}{d} = \left(\frac{\varepsilon_0\varepsilon_r A}{d}\right) F \tag{2.35}$$

If it is required to build such a capacitor with C as near to the theoretical value as possible, then the edge effects can be overcome by constructing one of the electrodes with a guard ring (Figure 2.8). The potential of the guard ring would be the same as that of the central electrode but has to be supplied from a separate source. Flux fringing occurs at the edges of the guard ring, but the flux density over the central electrode remains uniform.

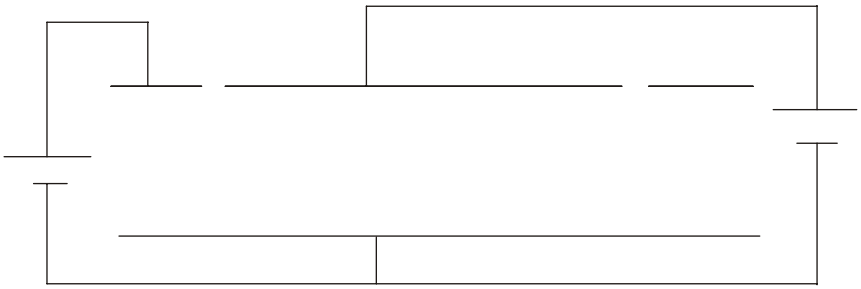


Figure 2.8 Guard ring of a parallel plate capacitor.

2.8.1.1 Two-plate tapered capacitor

We now consider a capacitor made up of two square metal plates (each of side b), mounted opposite to each other in such a way that their planes intersect in a line parallel to, and at a distance a from the nearer edge of each plate. The angle between the intersecting planes is α, the surrounding medium being air. We neglect the edge effects. Since the plates AB and CD are equipotential surfaces, the field lines would be arcs of circles with the centres at the point of intersection of the two plates, i.e. the point O. If the applied potential difference between the plates is V, then at a radius r [as shown in Figure 2.8(a)],

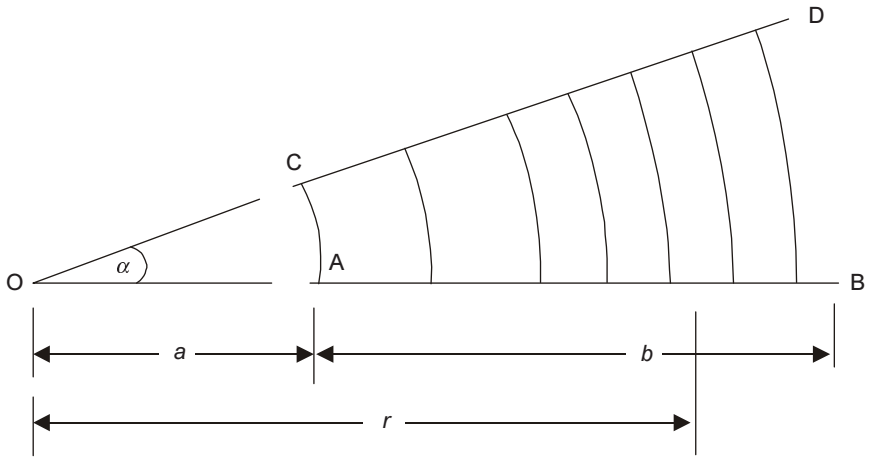


Figure 2.8(a) Two-plate tapered capacitor.

$E_\phi(r)$ in the region between the plates = $\dfrac{V}{\alpha r}$.

$$\therefore \qquad D_\phi(r) = \varepsilon_0 E_\phi(r) = \frac{\varepsilon_0 V}{\alpha r}$$

and

$$\frac{\varepsilon_0 V}{\alpha r} \quad [= \rho_S(r), \text{ by Gauss' theorem}]$$

where $\rho_S(r)$ is the surface charge at the radius r, on the plate surface.

$\therefore$ The total charge on the plate

$$Q = \iint_S \rho_S(r)\, dS$$

$$= b \int_{r=a}^{r=a+b} \frac{\varepsilon_0 V}{\alpha r}\, dr$$

$$= \left(\frac{b\varepsilon_0 V}{\alpha}\right) \ln\left(\frac{a+b}{a}\right)$$

$\therefore$

$$C = \frac{Q}{V} = \left(\frac{b\varepsilon_0}{\alpha}\right) \ln\left(1 + \frac{b}{a}\right) \tag{2.35a}$$

It should be noted that the lines of force in this problem are concentric arcs, the equipotentials are radial lines between the plates and the surface charge density on the equipotential plates is not a constant but a function of r.

2.8.2 Concentric Cylinders

This calculation is useful both for coaxial cables as well as for capacitors of concentric cylindrical design. Let the radii of the concentric cylindrical electrodes be a and b respectively (Figure 2.9), where $a < b$.

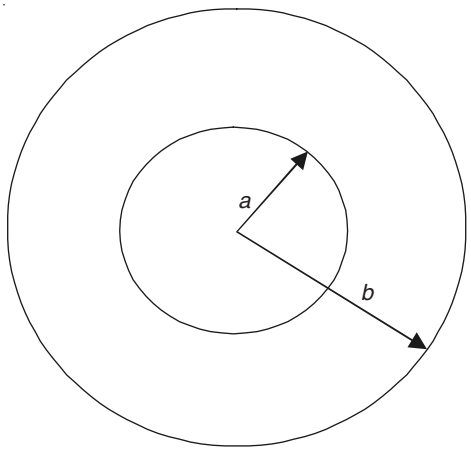


Figure 2.9 Concentric cylinders.

To avoid the edge effects we consider a system of infinite length axially, out of which a piece of unit length is taken. Let the charge on the inner cylinder be Q C/m.

At any radius r, $a \le r \le b$,

$$D = \frac{Q}{2\pi r \cdot 1} \tag{2.36}$$

from Gauss' theorem and considerations of symmetry.

$$\therefore \qquad E = \frac{Q}{2\pi\varepsilon r} \text{ V/m} \tag{2.37}$$

$\therefore$ The P.D. between the inner and the outer cylinders

$$= \int_a^b \mathbf{E} \cdot d\mathbf{r} = \int_a^b \frac{Q}{2\pi\varepsilon r}\, dr$$

$$= \frac{Q}{2\pi\varepsilon} \ln \frac{b}{a} \tag{2.38}$$

$$\therefore \qquad C = \frac{Q}{\text{P.D.}} = \frac{2\pi\varepsilon}{\ln\left(\frac{b}{a}\right)} \text{ F/m} \tag{2.39}$$

2.8.3 Parallel Circular Cylinders

Consider two infinitely long cylinders, each of radius a and carrying charges $\pm Q$ C/m (Figure 2.10). Let the distance between the centres of the cylinders be d. We know from the problem of the two parallel line charges [Eq. (1.49) of Section 1.6.4] that the potential at a point P due to such an arrangement is

$$V_P = \left(\frac{Q}{2\pi\varepsilon}\right) \ln\left(\frac{r_2}{r_1}\right) \tag{2.40}$$

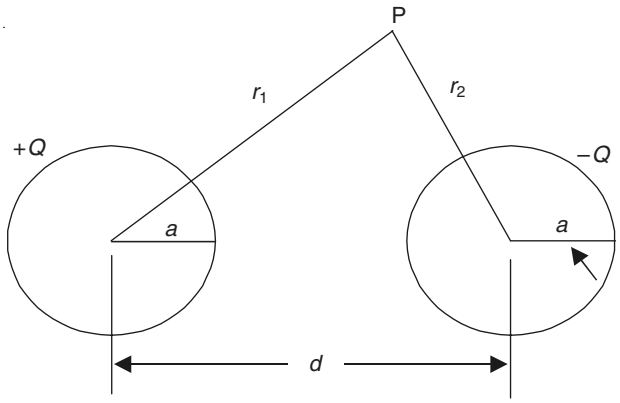


Figure 2.10 Parallel circular cylinders.

where the equipotentials were circles. Now when $r_1 << r_2$, these circles become more nearly centred on the centre of the left-hand line charge, and then become coincident with the equipotential surface of the cylinder of radius a if $a << d$.

$$\therefore \qquad V_1 \text{ on the left-hand cylinder} = \left(\frac{Q}{2\pi\varepsilon}\right) \ln\left(\frac{d}{a}\right) \tag{2.41}$$

and

$$V_2 \text{ on the right-hand cylinder} = \left(\frac{Q}{2\pi\varepsilon}\right)\ln\left(\frac{a}{d}\right)$$

$$= -\left(\frac{Q}{2\pi\varepsilon}\right)\ln\left(\frac{d}{a}\right) \tag{2.41a}$$

$$\therefore \quad \text{The P.D. between the two cylinders} = \left(\frac{Q}{\pi\varepsilon}\right)\ln\left(\frac{d}{a}\right) \tag{2.42}$$

$$\therefore \quad \text{Capacitance, } C = \frac{Q}{\text{P.D.}} = \frac{\pi\varepsilon}{\ln\left(\frac{d}{a}\right)} \tag{2.43}$$

Note that this problem cannot be solved if the cylinders of finite radii are replaced by infinitely thin line charges.

2.8.4 Wire and Parallel Plane

This problem is solved by the same technique as used in Section 2.8.3. So we again consider the arrangement of two parallel line charges (Figure 2.11) and have a look at the position of the point

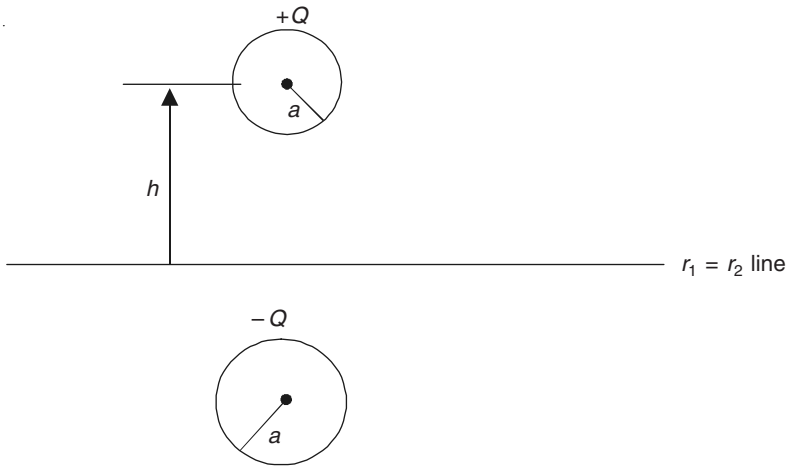


Figure 2.11 Wire and parallel conducting plane.

P in Figure 2.10, when $r_1 = r_2$.

From Eq. (2.40), it follows that $V_P = 0$ for $r_1 = r_2$. Hence we now insert a conducting plane through this line parallel to the line charge. This would have no effect on the field as the plane is at zero potential. If the wire has the radius a, and the distance between the wire and the plane is h, then

$$V_1 \text{ on the wire} = \left(\frac{Q}{2\pi\varepsilon}\right)\ln\left(\frac{2h}{a}\right) \tag{2.44}$$

$$V_2 \text{ on the plane} = 0$$

$$\therefore \quad \text{P.D.} = V_1 - V_2 = \left(\frac{Q}{2\pi\varepsilon}\right)\ln\left(\frac{2h}{a}\right) \tag{2.45}$$

$\therefore$ Capacitance per unit length of the wire, i.e.

$$C/\text{metre} = \frac{Q}{V_1 - V_2} = \frac{2\pi\varepsilon}{\ln\left(\frac{2h}{a}\right)} \text{ F/m} \tag{2.46}$$

The method used here for solving this problem, treats the second wire of the previous problem to be equivalent to 'an image' of the first wire in the conducting plane. This method is called the 'method of images' and is widely used for solving problems in electrostatics as well as in magnetostatics. We shall discuss the details and the mathematical background of this method later, though a brief note is added below.

2.8.5 An Introductory Note on the Method of Images

In general terms, the method of electric images implies determining an imaginary distribution of charges inside a conducting body of an electrostatic system, which would produce outside the body exactly the same field as that produced by the actual free charges over its surface.

The basic requirement of the method of images is that the effect of the boundaries shall be correctly represented by the images. The field of the image charges within this region must be identical with that of the induced charges (which the image charges are replacing).

The necessary condition to determine the uniqueness of the field is that the value of the tangential or the normal **E** field be specified at every point of the surface.

2.8.6 Capacitance between Two Spheres of Equal Diameter

Let the spheres be of radius a and the distance between the centres of the two spheres be d, where $d >> 2a$ (see Figure 2.12). Consider a point P such that its distances from the centres of the spheres are r_1 and r_2, respectively, as shown in Figure 2.12.

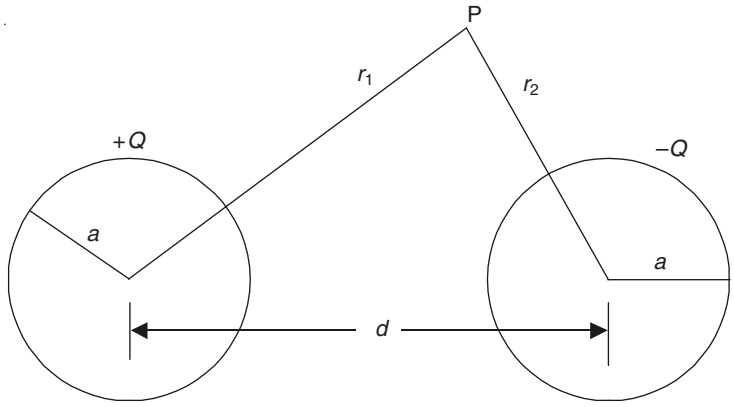


Figure 2.12 Two spheres of equal diameters.

Now we know from Eq. (1.37) that the potential at a point P due to an isolated charged sphere whose centre is distant r from it, is given by

$$V_P = \frac{Q}{4\pi\varepsilon r} \tag{2.47}$$

$\therefore$ When there are two charged spheres with charges $+Q$ and $-Q$ as in this case, then

$$V_P = \frac{1}{4\pi\varepsilon}\left(\frac{Q}{r_1} - \frac{Q}{r_2}\right) \tag{2.48}$$

Note that as $r_1 << r_2$, $V_P \to Q/4\pi\varepsilon r_1$, i.e. the equipotentials approach to the spherical surfaces centred on the point charge.

Coming back to the problem of the two spheres (Figure 2.12),

$$V_1 \text{ on the left-hand sphere } = \left(\frac{Q}{4\pi\varepsilon}\right)\left(\frac{1}{d} - \frac{1}{a}\right) \tag{2.49}$$

$$V_2 \text{ on the right-hand sphere } = \left(\frac{Q}{4\pi\varepsilon}\right)\left(\frac{1}{a} - \frac{1}{d}\right)$$

$\therefore$ The P.D. between the two spheres $= V = V_2 - V_1$

$$= \left(\frac{Q}{2\pi\varepsilon}\right)\left(\frac{1}{a} - \frac{1}{d}\right) \tag{2.50}$$

$\therefore$ Capacitance, $$C = \frac{Q}{\text{P.D.}} = \frac{2\pi\varepsilon}{\frac{1}{a} - \frac{1}{d}} \text{ F} \tag{2.51}$$

2.8.7 Capacitance between a Sphere and a Conducting Plane

Once again this is a problem similar to that mentioned in Section 2.8.4. So the 'method of images' can be applied to solve this problem. It can be easily shown that if a sphere of radius a whose centre is at distance d from the conducting plane, then the capacitance of the system is given by

$$C = \frac{4\pi\varepsilon}{\frac{1}{a} - \frac{1}{d}} \text{ F} \tag{2.52}$$

2.8.8 Capacitances in Parallel and in Series

If a number of capacitors of capacitances C_1, C_2, C_3, ... are all connected in parallel, the potential difference between them will be the same, say V (Figure 2.13).

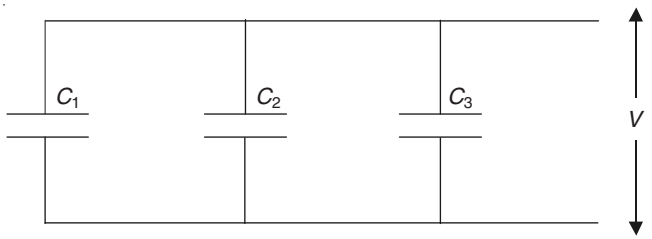


Figure 2.13 Capacitors in parallel.

The charges on these capacitors would be C_1V, C_2V, C_3V, ... Thus the total charge Q of the system will be given by $Q = (C_1 + C_2 + C_3 + \cdots)V$. Therefore, the effective capacitance of the system is

$$C = \frac{Q}{V} = C_1 + C_2 + C_3 + \cdots \tag{2.53}$$

Thus the system behaves as if it is a single capacitor of capacitance $(C_1 + C_2 + C_3 + \cdots)$.

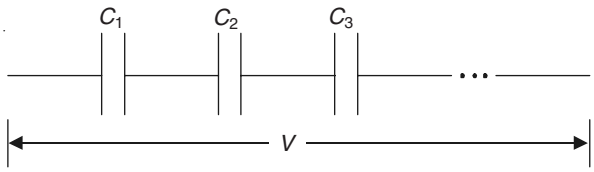


Figure 2.14 Capacitors in series.

Let us now consider the system in which the capacitors are connected in series (Figure 2.14). The capacitors are initially all discharged. Initially, if negative charge $-Q$ is removed from the electrode a_1, it gets charged by $+Q$. This charges the electrode b_1 by $-Q$, which in turn charges a_2 by $+Q$, and so on. Thus it is seen that each capacitor acquires the same charge. So from the corresponding P.D. of each capacitor, the total P.D. of the system would be:

$$V = \frac{Q}{C_1} + \frac{Q}{C_2} + \frac{Q}{C_3} + \cdots$$

$$\therefore \quad \frac{Q}{V} = \frac{1}{\frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \cdots}$$

That is, it is effectively a single capacitor of capacitance C given by

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \cdots \tag{2.54}$$

2.9 FIELD IN A REGION CONTAINING TWO DIELECTRIC MATERIALS: BOUNDARY CONDITIONS IN ELECTROSTATICS

So far in all the situations considered, we have tacitly assumed that in the intervening space

between the conductors, there is a single dielectric, that is, while calculating the capacitances we have taken for granted that the vacuum between the conductors has been replaced by a single uniform insulating material. This assumption has been the basis for the definition of dielectric, and also for the generalization of the inverse square law of forces. We shall generalize further to consider the situation when the medium is not uniform, that is, there is more than one dielectric present. So we shall now see as to what happens on a surface Σ where one dielectric (ε_1) meets

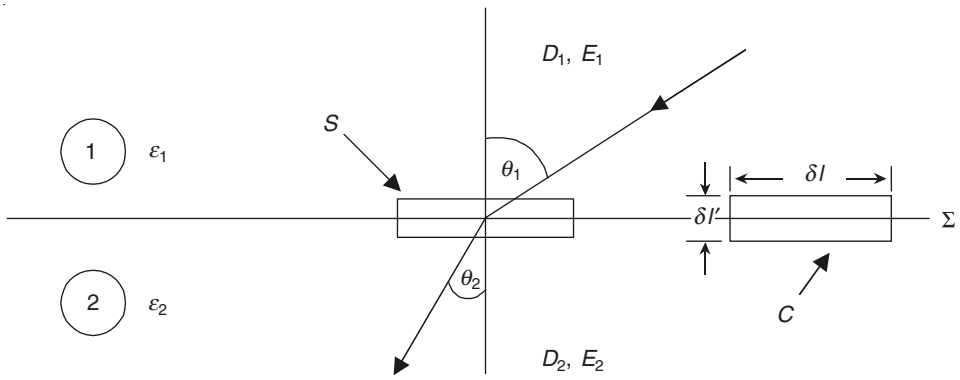


Figure 2.15 Surface of discontinuity.

another dielectric (ε_2). (See Figure 2.15.) Such a surface is a surface of discontinuity. On Σ, consider a small coin-shaped Gaussian surface S with cross-sectional area δS as shown in Figure 2.15, enclosing a small portion of Σ. Then for this closed surface,

$$(D_{1n} - D_{2n})\delta S + \text{contribution from the edge of } S = \text{enclosed charge} = \rho_S\, \delta S \tag{2.55}$$

where D_{1n} and D_{2n} are the normal components of $\mathbf{D}_1$ and $\mathbf{D}_2$, respectively, on this interface from the two sides, and ρ_S is the surface charge density on Σ. In the limit, the contribution from the edges of $S \to 0$. Also, $D_{1n} = D_1 \cos \theta_1$ and $D_{2n} = D_2 \cos \theta_2$.

$$\therefore \qquad D_{1n} - D_{2n} = \rho_S \tag{2.56}$$

Consider again a rectangular circuit C across the surface Σ, as shown, and take the line integral of $\mathbf{E}$ along this loop.

$$\therefore \qquad (E_{1t} - E_{2t}) + \text{edge contributions along } \delta l' = 0 \tag{2.57}$$

This follows from the condition that $\oint \mathbf{E} \cdot d\mathbf{l} = 0$,

$$\because \qquad \mathbf{E} = -\text{ grad } V$$

In the limit the edge contributions will again $\to 0$.

$$\therefore \qquad E_{1t} - E_{2t} = 0 \tag{2.58}$$

where E_{1t} and E_{2t} are the tangential components of $\mathbf{E}_1$ and $\mathbf{E}_2$, respectively, and hence $E_{1t} = E_1 \sin \theta_1$ and $E_{2t} = E_2 \sin \theta_2$. So the conclusions are:

(i) the normal component of $\mathbf{D}$ is continuous if there is no charge across the surface of discontinuity; and
(ii) the tangential component of $\mathbf{E}$ is continuous across this surface.

Some special cases

(a) *When there is no surface charge on the interface between the two dielectrics* (Figure 2.16)

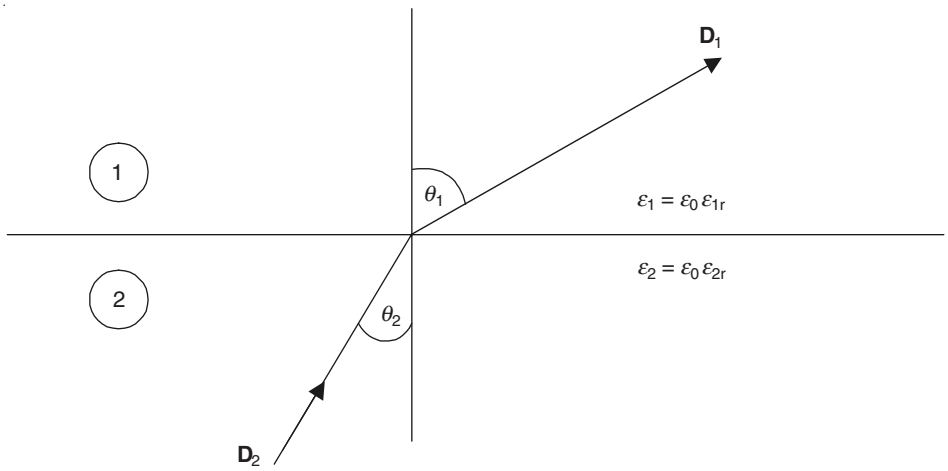


Figure 2.16 No surface charge on the interface.

For this case, $D_{1n} = D_{2n}$ and $E_{1t} = E_{2t}$

Also, $\mathbf{D}_1 = \varepsilon_1 \mathbf{E}_1 = \varepsilon_0 \varepsilon_{1r} \mathbf{E}_1$ and $\mathbf{D}_2 = \varepsilon_2 \mathbf{E}_2 = \varepsilon_0 \varepsilon_{2r} \mathbf{E}_2$

$$\therefore \quad \frac{D_{1n}}{E_{1t}} = \frac{D_{1n}}{\frac{D_{1t}}{\varepsilon_0 \varepsilon_{1r}}}$$

$$= \frac{D_{2n}}{E_{2t}} = \frac{D_{2n}}{\frac{D_{2t}}{\varepsilon_0 \varepsilon_{2r}}} \tag{2.59}$$

Also,

$$\frac{D_{1n}}{D_{1t}} = \tan \theta_1 \quad \text{and} \quad \frac{D_{2n}}{D_{2t}} = \tan \theta_2$$

$\therefore$ From Eq. (2.59), we get

$$\frac{\varepsilon_{1r}}{\tan \theta_1} = \frac{\varepsilon_{2r}}{\tan \theta_2} \quad \text{or} \quad \frac{\tan \theta_1}{\tan \theta_2} = \frac{\varepsilon_{1r}}{\varepsilon_{2r}} \tag{2.60}$$

(b) *On the interface* (*carrying a surface charge density* ρ_S) *of a conductor and a dielectric medium* (Figure 2.17).

Consider a coin-shaped Gaussian surface S on this interface. For this closed surface,

$$\oiint_S \mathbf{D} \cdot d\mathbf{S} = \text{enclosed charge} = \rho_S \, dS$$

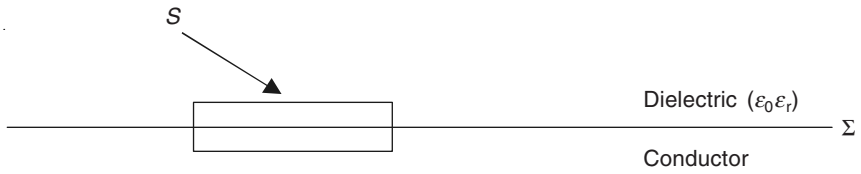


Figure 2.17 Conductor and dielectric.

where dS is the cross-sectional area of the Gaussian surface S. Since there is surface charge on the outer face only of the conducting medium,

$$D_n = 0, \text{ in the conductor} \tag{2.61}$$

It is only in the dielectric medium, that the normal component of **D** is nonzero.

$$\therefore \quad D_n = \rho_S \quad \text{or} \quad E_n = \frac{\rho_S}{\varepsilon_0 \varepsilon_r} \tag{2.62}$$

$$\left(= \frac{\rho_S}{\varepsilon_0}, \text{ if the medium is vacuo or air.}\right)$$

2.10 CAPACITORS OF MIXED DIELECTRICS AND OF COMPLEX SHAPES

2.10.1 Parallel Plate Capacitor with Mixed Dielectrics

The spacing between the two plates is $(d_1 + d_2)$, and the thickness of the dielectric with the permittivity ε_1 $(= \varepsilon_0\varepsilon_{1r})$ is d_1, and that of the dielectric with the permittivity ε_2 $(= \varepsilon_0\varepsilon_{2r})$ is d_2 (see Figure 2.18).

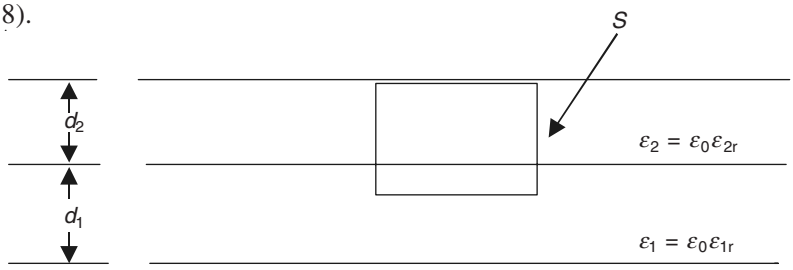


Figure 2.18 Parallel plate capacitor with mixed dielectrics.

By considering a cylindrical Gaussian surface S with one end inside one of the plates and the other in the dielectric, we obtain

$$\oiint_S \mathbf{D} \cdot d\mathbf{S} = \rho_S dS \tag{2.63}$$

where dS is the cross-sectional area and ρ_S is the surface charge density on the plate. This is because there are no contributions to the flux from the sides of the cylindrical surface if the edge effects are neglected (i.e. assuming the capacitor to extend to infinity).

$$\therefore \quad D = \rho_S \tag{2.64}$$

$$\therefore \qquad E_1 = \frac{\rho_S}{\varepsilon_0 \varepsilon_{1r}} \qquad \text{and} \qquad E_2 = \frac{\rho_S}{\varepsilon_0 \varepsilon_{2r}} \tag{2.65}$$

That is, **E** is greater where the permittivity is smaller. Therefore, the potential difference (= V) between the plates is given by

$$V = E_1 d_1 + E_2 d_2$$

$$= \left(\frac{\rho_S}{\varepsilon_0}\right)\left(\frac{d_1}{\varepsilon_{1r}} + \frac{d_2}{\varepsilon_{2r}}\right) \tag{2.66}$$

If the area of the plates is A, and the edge effects are neglected, then the capacitance C is given by

$$C = \rho_S \frac{A}{V} = \varepsilon_0 \frac{A}{\dfrac{d_1}{\varepsilon_{1r}} + \dfrac{d_2}{\varepsilon_{2r}}} \tag{2.67}$$

This should be compared with Eq. (2.35).

2.10.2 Concentric Cylinders with Mixed Dielectrics

Another type of mixed dielectric capacitor is the one with concentric cylinders. The radii of the electrodes are a and c (Figure 2.19), where $a < c$.

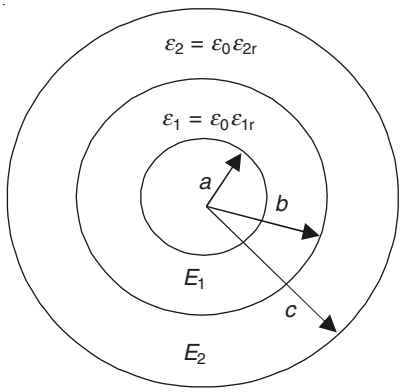


Figure 2.19 Concentric cylinders with mixed dielectrics.

A dielectric of permittivity $\varepsilon_1 (= \varepsilon_0 \varepsilon_{1r})$ extends from the inner electrode to the radius b (where $a < b < c$), and the second dielectric of permittivity $\varepsilon_2 (= \varepsilon_0 \varepsilon_{2r})$ extends from the radius b to the outer electrode at the radius c. The surface of discontinuity is the radial interface at $r = b$. As in the problem of Section 2.8.2, the electric flux density at any radius r ($a < r < c$) is given by

$$D = \frac{Q}{2\pi r \cdot 1} \tag{2.68}$$

where Q is the charge per unit length of the electrode.

$$\therefore \qquad E_1 = \frac{Q}{2\pi \varepsilon_0 \varepsilon_{1r} r} \text{ V/m} \quad \text{and} \quad E_2 = \frac{Q}{2\pi \varepsilon_0 \varepsilon_{2r} r} \text{ V/m} \tag{2.69}$$

Hence the potential difference between the two electrodes is

$$= \int_a^b \mathbf{E}_1 \cdot d\mathbf{r} + \int_b^c \mathbf{E}_2 \cdot d\mathbf{r}$$

as only the radial component of **E** exists.

$$\therefore \quad \text{P.D.} = \left(\frac{Q}{2\pi\varepsilon_0}\right)\left[\left(\frac{1}{\varepsilon_{1r}}\right)\int_a^b \frac{dr}{r} + \left(\frac{1}{\varepsilon_{2r}}\right)\int_b^c \frac{dr}{r}\right]$$

$$= \left(\frac{Q}{2\pi\varepsilon_0}\right)\left[\left(\frac{1}{\varepsilon_{1r}}\right)\ln\left(\frac{b}{a}\right) + \left(\frac{1}{\varepsilon_{2r}}\right)\ln\left(\frac{c}{b}\right)\right] \tag{2.70}$$

$$\therefore \quad \text{Capacitance,} \quad C = \frac{Q}{\text{P.D.}}$$

$$= \frac{2\pi\varepsilon_0}{\left(\frac{1}{\varepsilon_{1r}}\right)\ln\left(\frac{b}{a}\right) + \left(\frac{1}{\varepsilon_{2r}}\right)\ln\left(\frac{c}{b}\right)} \tag{2.71}$$

This should be compared with Eq. (2.39). It will be noticed that both these types, i.e. parallel plate and concentric cylinders, with mixed dielectrics behave like two single dielectric capacitors connected in series.

$$\therefore \quad \frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2}$$

2.10.3 Concentric Spheres with Single Dielectric

This problem and the subsequent problem with mixed dielectrics have been solved by a different method. It has been shown in Section 1.5.2. that the potential field in free space is Laplacian in nature. This has been proved for free space and it follows directly that this equation holds for dielectric materials as well. We shall now use the solution of this equation, i.e. Laplace's equation, to evaluate the capacitance of concentric spheres.

The two concentric spheres have the radii a, b ($a < b$), respectively, and the potentials of the electrodes are V_a and V_b, respectively (Figure 2.20).

Since the potential distribution V in the intervening dielectric (carrying no charge) satisfies the Laplace's equation, we have $\nabla^2 V = 0$. Since the geometry of this capacitor is spherical, we use the spherical polar coordinate system. In this system,

$$\nabla^2 V \equiv \frac{1}{r^2}\left[\frac{\partial}{\partial r}\left(r^2 \frac{\partial V}{\partial r}\right)\right] + \frac{1}{r^2 \sin\theta}\left[\frac{\partial}{\partial \theta}\left(\sin\theta \frac{\partial V}{\partial \theta}\right)\right] + \frac{1}{r^2 \sin^2\theta}\left[\left(\frac{\partial^2 V}{\partial^2 \phi}\right)\right] = 0 \tag{2.72}$$

Because of the symmetry of the problem,

$$\frac{\partial V}{\partial \theta} = 0 \qquad \text{and} \qquad \frac{\partial V}{\partial \phi} = 0$$

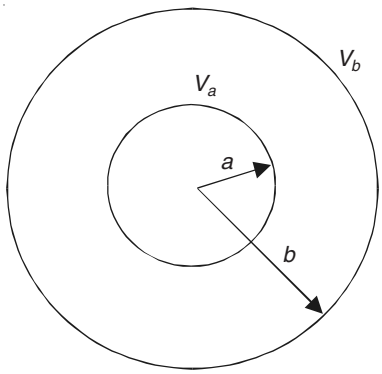


Figure 2.20 Concentric spheres.

$\therefore$ Equation (2.72) simplifies to

$$\frac{1}{r^2}\left[\frac{\partial}{\partial r}\left(r^2\frac{\partial V}{\partial r}\right)\right]=0 \qquad \text{or} \qquad \frac{\partial}{\partial r}\left(r^2\frac{\partial V}{\partial r}\right)=0$$

Integrating, $r^2\dfrac{\partial V}{\partial r} = A$, where A is the constant of integration. We have, integrating again,

$$V = A\int\frac{dr}{r^2}+B=-\frac{A}{r}+B$$

To evaluate the constants of integration A and B, we use the following boundary conditions:

$$\text{At } r=a,\ V=V_a=-\frac{A}{a}+B \qquad \text{and} \qquad \text{at } r=b,\ V=V_b=-\frac{A}{b}+B$$

$\therefore$
$$A=\frac{ab(V_b-V_a)}{b-a}, \quad B=\frac{bV_b-aV_a}{b-a}$$

$\therefore$
$$V=-\frac{ab(V_b-V_a)}{b-a}\frac{1}{r}+\frac{bV_b-aV_a}{b-a} \tag{2.73}$$

From the potential distribution, the electric field intensity **E** can be evaluated by using

$$\mathbf{E}=-\text{ grad } V=-\left[\mathbf{i}_r\left(\frac{\partial V}{\partial r}\right)+\mathbf{i}_\theta\left(\frac{1}{r}\right)\left(\frac{\partial V}{\partial \theta}\right)+\mathbf{i}_\phi\left(\frac{1}{r\sin\theta}\right)\left(\frac{\partial V}{\partial \phi}\right)\right]$$

in the spherical polar coordinate system.

$\therefore$
$$\mathbf{E}=-\mathbf{i}_r\left(\frac{\partial V}{\partial r}\right), \text{ in this case because of symmetry.}$$

$$=-\left[\frac{ab(V_b-V_a)}{b-a}\right]\left(\frac{1}{r^2}\right)\mathbf{i}_r$$

$$\therefore \qquad \mathbf{D} = \varepsilon\mathbf{E} = -\mathbf{i}_r \left[\frac{\varepsilon ab(V_b - V_a)}{b-a}\right]\left(\frac{1}{r^2}\right)$$

Hence on the surface of the inner electrode (i.e. at $r = a$)

$$D_a = \left[\frac{\varepsilon ab\,(V_b - V_a)}{b-a}\right]\left(\frac{1}{a^2}\right) \tag{2.74}$$

This is the charge per unit area of the surface of the electrode, and the total charge will be $4\pi a^2$ times this quantity.

$$\therefore \qquad \text{Capacitance, } C = \frac{\text{charge}}{\text{P.D.}} = \frac{4\pi\varepsilon ab}{b-a}$$

$$= \frac{4\pi\varepsilon_0\varepsilon_\text{r}}{\frac{1}{a} - \frac{1}{b}} \tag{2.75}$$

2.10.4 Concentric Spheres with Mixed Dielectrics

This problem can also be solved by the same method, except that on the interface between the two dielectrics, the relevant interface continuity conditions have to be used to evaluate the constants of integration. The dimensions and the physical parameters are shown in Figure 2.21.

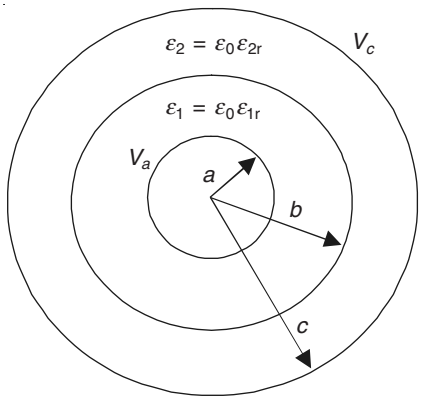


Figure 2.21 Concentric spheres with mixed dielectrics.

As before, the potential distributions in the dielectric regions are:

$$V_1 = -\frac{A_1}{r} + B_1, \text{ in } \varepsilon_1 \quad \text{and} \quad V_2 = -\frac{A_2}{r} + B_2, \text{ in } \varepsilon_2 \tag{2.76}$$

Then, from $\mathbf{E} = -\text{grad } V = -\mathbf{i}_\text{r}\dfrac{\partial V}{\partial r}$, we get

$$\mathbf{E}_1 = -\mathbf{i}_\text{r}\frac{A_1}{r^2},\ a < r < b; \quad \mathbf{E}_2 = -\mathbf{i}_\text{r}\frac{A_2}{r^2},\ b < r < c$$

$$\therefore \qquad \mathbf{D}_1 = -\mathbf{i}_r\varepsilon_0\varepsilon_{1\text{r}}\frac{A_1}{r^2},\ a < r < b \quad \text{and} \quad \mathbf{D}_2 = -\mathbf{i}_r\varepsilon_0\varepsilon_{2\text{r}}\frac{A_2}{r^2},\ b < r < c$$

There are four unknowns and hence we need four boundary conditions to evaluate them. They are:

$$\text{On } r = a,\ V_1 = V_a = -\frac{A_1}{a} + B_1$$

$$\text{On } r = c,\ V_2 = V_c = -\frac{A_2}{c} + B_2$$

$$\text{On } r = b,\ V_1 = V_2, \quad \text{i.e.} \quad -\frac{A_1}{b} + B_1 = -\frac{A_2}{b} + B_2$$

$$\text{On } r = b,\ D_n \text{ is continuous, so } D_{n1} = D_{n2} \qquad \text{or} \qquad D_{r1} = D_{r2};\ \text{i.e. } \varepsilon_0\varepsilon_{1r}\frac{A_1}{b^2} = \varepsilon_0\varepsilon_{2r}\frac{A_2}{b^2}$$

From these boundary conditions, we get **D** on the surface of the inner electrode, i.e. at $r = a$,

$$D_1 = \frac{\varepsilon_0 (V_c - V_a)}{\left\{\frac{1}{\varepsilon_{2r}}\left(\frac{1}{b} - \frac{1}{c}\right) + \frac{1}{\varepsilon_{1r}}\left(\frac{1}{a} - \frac{1}{b}\right)\right\}}\frac{1}{a^2}$$

Therefore, the total charge on the inner sphere will be $4\pi a^2$ times this quantity.

$$\therefore \qquad \text{Capacitance, } C = \frac{\text{charge}}{V_c - V_a} = \frac{4\pi\varepsilon_0}{\frac{1}{\varepsilon_{2r}}\left(\frac{1}{b} - \frac{1}{c}\right) + \frac{1}{\varepsilon_{1r}}\left(\frac{1}{a} - \frac{1}{b}\right)} \tag{2.77}$$

So, as for the other two mixed dielectric capacitors, this one is also effectively like the two single dielectric capacitors connected in series.

These two problems can also be solved directly by integration which is a much shorter and a more direct method. This method is now given below:

Assuming a charge Q on the inner sphere, we have

$$E_1 = \frac{Q}{4\pi\varepsilon_0\varepsilon_{1r}r^2}$$

$$\therefore \qquad V_1 = \int_a^b \mathbf{E} \cdot d\mathbf{r} = \frac{Q}{4\pi\varepsilon_0\varepsilon_{1r}}\int_a^b \frac{dr}{r^2}$$

$$= \frac{Q}{4\pi\varepsilon_0\varepsilon_{1r}}\left(\frac{1}{a} - \frac{1}{b}\right)$$

= P.D. across the first dielectric ε_1

Similarly,

$$V_2 = \int_b^c \frac{Q}{4\pi\varepsilon_0\varepsilon_{2r}}\left(\frac{dr}{r^2}\right)$$

$$= \frac{Q}{4\pi\varepsilon_0\varepsilon_{2r}}\left(\frac{1}{b} - \frac{1}{c}\right)$$

$$\therefore \quad \text{Total P.D.} = V_1 + V_2 \quad \text{and} \quad \text{Capacitance } C = \frac{Q}{V_1 + V_2}$$

$$= \frac{4\pi\varepsilon_0}{\frac{1}{\varepsilon_{1r}}\left(\frac{1}{a} - \frac{1}{b}\right) + \frac{1}{\varepsilon_{2r}}\left(\frac{1}{b} - \frac{1}{c}\right)} \tag{2.77a}$$

2.10.5 Capacitance of N Conductors

We now generalize the concept of capacitance from two conductors to N conductors. They are all insulated and initially at zero potential. The earth is also at zero potential, treated as an infinite charge reservoir of unvarying potential.

Let a charge q_1 be conveyed from the earth to the conductor 1. The potential of conductor 1 will be raised and other conductors will be at intermediate values. Now,

$$V_1' = p_{11}q_1, \; V_2' = p_{21}q_1, \; ..., \; V_N' = p_{N1}q_1 \tag{2.78a}$$

where p_{11}, p_{21}, ..., p_{N1} are constants.

Similarly, transferring a charge q_2 from the earth to conductor 2, will set up potentials as

$$V_1'' = p_{21}q_2, \; V_2'' = p_{22}q_2, \; ..., \; V_N'' = p_{N2}q_2 \tag{2.78b}$$

By the principle of superimposition, when the charges q_1, q_2, ..., q_N are moved to N conductors respectively, the final potential distribution will be:

$$\begin{aligned} V_1 &= p_{11}q_1 + p_{12}q_2 + \cdots + p_{1N}q_N \\ V_2 &= p_{21}q_1 + p_{22}q_2 + \cdots + p_{2N}q_1 \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ V_N &= p_{N1}q_1 + p_{N2}q_2 + \cdots + p_{NN}q_N \end{aligned} \tag{2.79}$$

The constants p_{ij} are called the coefficients of potential. These N equations can be solved for charges in terms of potential, i.e.

$$\begin{aligned} q_1 &= c_{11}V_1 + c_{12}V_2 + \cdots + c_{1N}V_N \\ q_2 &= c_{21}V_1 + c_{22}V_2 + \cdots + c_{2N}V_N \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ q_N &= c_{N1}V_1 + c_{N2}V_2 + \cdots + c_{NN}V_N \end{aligned} \tag{2.80}$$

These equations are the counterparts of the definition of C, i.e. $q = CV$. Here, c_{11}, c_{12}, ... c_{21}, ... are constants, dependent on the shape and location of the bodies, i.e. c_{11} signifies the charge on conductor 1 when it is raised to a potential of 1 V while others are earthed (i.e. kept at zero potential) etc. So, c_{11}, c_{22}, ..., c_{NN} are called the coefficients of capacitance. The coefficient c_{21} is the charge induced on the conductor 2 when the conductor 1 is raised to the potential of 1 V and the other conductors are earthed. The other constants, c_{12}, c_{13}, ... are called the coefficients of induction. The coefficients of potential and induction are not all different but equal in pairs i.e. $p_{12} = p_{21}$, $c_{12} = c_{21}$, ..., and so on.

Proof: Let the charge q_1 be gradually placed on the conductor 1 in small steps of δq. When the

charge reaches the value q, its potential is $p_{11}q$. Therefore, the work required to bring up the next increment δq is $p_{11}q\delta q$. Therefore the work required to bring up the whole charge q_1 is given by

$$\int_0^{q_1} p_{11} q\, dq = \frac{p_{11} q_1^2}{2}$$

Similarly, q_2 is to be brought up gradually to the conductor 2 in small steps of δq.
When the charge on the conductor 2 is q, its potential is $(p_{21}q_1 + p_{22}q)$.
Therefore the work done in bringing up the next increment of charge = $(p_{21}q_1 + p_{22}q)\delta q$.
Therefore the total work done in raising the charge of the conductor 2 to q_2 is

$$\int_0^{q_2} (p_{21}q_1 + p_{22}q)\, dq = p_{21}q_1q_2 + \frac{p_{22}q_2^2}{2} \tag{2.81}$$

Hence the total work done in bringing up the charges q_1 and q_2

$$= \frac{p_{11}q_1^2}{2} + p_{21}q_1q_2 + \frac{p_{22}q_2^2}{2} \tag{2.82}$$

If the charging order had been reversed, then the work done

$$= \frac{p_{22}q_2^2}{2} + p_{12}q_2q_1 + \frac{p_{11}q_1^2}{2} \tag{2.83}$$

Since the work done is same in both the cases, we have

$$p_{12} = p_{21} \tag{2.84}$$

From this it can be shown algebraically, that

$$c_{12} = c_{21} \tag{2.85}$$

and so on.

The first equation of the system can be rewritten as

$$\begin{aligned} q_1 &= c_{11}V_1 + c_{12}V_2 + c_{13}V_3 + \cdots + c_{1N}V_N \\ &= (c_{11} + c_{12} + \cdots + c_{1N})V_1 - c_{12}(V_1 - V_2) - c_{13}(V_1 - V_3) - \cdots - c_{1N}(V_1 - V_N) \end{aligned}$$

or

$$q_1 = C_1V_1 + C_{12}(V_1 - V_2) + C_{13}(V_1 - V_3) + \cdots + C_{1N}(V_1 - V_N) \tag{2.86}$$

where

$$\begin{aligned} C_1 &= c_{11} + c_{12} + \cdots + c_{1N} \\ C_{12} &= -c_{12},\ C_{13} = -c_{13}, \text{ and so on} \end{aligned} \tag{2.87}$$

Similarly for q_2, q_3, q_4, ..., q_N.

Thus for the system, C_1, C_2, C_3, ..., C_N are the capacitances between the bodies and the earth; and C_{12}, C_{13}, ... are the capacitances between the bodies taken in pairs.

This technique can be used for calculating the capacitances associated with practical problems such as that of, say, a three-phase transmission line, or for analyzing the problem of electrostatic screening.

PROBLEMS

2.1 A capacitor is formed of tin-foil sheets applied to the two faces of a glass plate of thickness 0.4 cm and relative permittivity 6. Very thin layers of air are trapped between the foil and glass. Given that the air becomes ionized when $E = 3 \times 10^6$ V/m, find *approximately* the potential difference at which the ionization will start in the capacitor.

Ans.: $V_b = 2$ kV

2.2 A pair of wires of radius a are held at a distance d apart, and each is at a distance h from a conducting plane. Prove that the capacitance between the two wires, connected in parrallel, and the plane is $4\pi\varepsilon_0/[\ln\{2hd'/(ad)\}]$ where $d'^2 = d^2 + 4h^2$, $a << d$ and h.

2.3 Two long wires, each of radius a, are held at a height h above an earthed conducting plane, paralled to each other and at a distance b apart. One wire is raised to a potential V_1 with respect to the plane, the other being insulated. Prove that the potential taken up by this wire is $V_1\{\ln (2h'/b)\}/\{\ln (2h/a)\}$, where $h'^2 = h^2 + b^2/4$.

2.4 Each of the two dielectrics (of relative permittivities ε_{r1} and ε_{r2} respectively) occupies one-half the volume of the annular space between the electrodes of a cylindrical capacitor, such that the interface plane between the dielectrics is a $r-z$ plane. Show that the two dielectrics act like a single dielectric having the average relative permittivity.

2.5 A parallel plate capacitor with free space between the electrodes is connected to a constant voltage source. If the plates are moved apart from the separation d to $2d$, keeping the potential difference between them unchanged, what will be the change in **D**? On the other hand, if the plates are brought closer together from d to $d/2$ with a dielectric of relative permittivity $\varepsilon_r = 3$, while maintaining the charges on the plates at the same value, what will be the changes in the P.D.?

Ans.: (i) $\mathbf{D}_2 = \mathbf{D}_1/2$ (ii) $V_2 = V_1/6$

3 Energy and Mechanical Forces in Electrostatic Fields

3.1 ELECTROSTATIC FORCES

The first thing to realize about the forces of electrostatic origin is that they are quite small compared with the forces of, say, magnetic fields. Such forces cannot be used to drive motors, but can be used only to operate instruments. So the present discussion of mechanical forces due to electrostatic fields would be a relatively short one.

3.2 ENERGY OF A SYSTEM OF CHARGED CONDUCTORS

The mechanical forces on charged conducting bodies in an electrostatic field can be obtained by adding up the forces on each pair of conductors, which in turn can be evaluated by the inverse square law. However the forces are vectors and the resultant force would be the vector sum of all the component forces. Hence it is more convenient to calculate the energy of the system—a scalar quantity—from which the forces can be derived.

The energy of a system of charged conductors is the work done, which is required to bring them to a charged state from a standard state—which is taken as a state of complete discharge. Let us consider a group of N conductors. When the charges Q_1, Q_2, Q_3, ..., Q_N have been conveyed from the earth and deposited on these N conductors, respectively, their potentials will be raised to V_1, V_2 , V_3, ..., V_N. Now we have to evaluate the energy of the system. So, let us suppose that the energy is being transferred in small steps, i.e. at a given time when the fraction kQ of the final energy Q (where $0 \le k \le 1$) has been transferred, the next fraction is δk. Thus the work done to bring up this next step is: (At this stage the charges are kQ_1, kQ_2, kQ_3, ..., kQ_N and the corresponding potentials are kV_1, kV_2, kV_3, ..., kV_N.)

$$\delta W = (\delta k Q_1)(kV_1) + (\delta k Q_2)(kV_2) + \cdots + (\delta k Q_N)(kV_N)$$
$$= k\delta k(Q_1V_1 + Q_2V_2 + \cdots + Q_NV_N).$$

$\therefore$ Work required for the whole charging process is

$$W = \int_0^1 k\,\delta k(Q_1V_1 + Q_2V_2 + \cdots + Q_NV_N)$$

$$= \frac{1}{2}(Q_1V_1 + Q_2V_2 + \ldots + Q_NV_N)$$

$$= \frac{1}{2}\Sigma(QV) \tag{3.1}$$

This is the energy expressed in joules.

$\therefore$ The energy of a charged capacitor $= \frac{QV}{2} = \frac{CV^2}{2}$

$$= \frac{Q^2}{2C} \tag{3.2}$$

3.3 ENERGY STORED IN THE ELECTRIC FIELD

When the energy in a group of charged conductors is expressed in the form $W = QV/2$, it is being associated with the charged conductors. This is the potential energy of the system, being associated with the relative positions of the conductors. However it is possible to express this energy in terms of the field quantities **E** and **D,** i.e. as if the energy is stored in the electric field. For instance, consider parts of two conductors in the system shown in Figure 3.1.

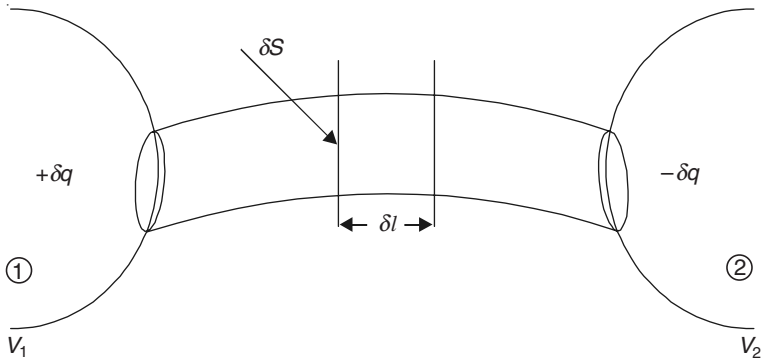


Figure 3.1 Energy in terms of field quantities.

Let the charge δq be on the conductor 1 and its counterpart $-\delta q$ on the other end of the flux tube on the conductor 2. (*Note:* Such a flux tube connecting the charge $+\delta q$ on one conductor to $-\delta q$ on the other conductor is called a 'Faraday tube'.) If the potentials of these two conductors are V_1 and V_2, respectively, then the contribution of these two charges to the energy is

$$\delta W = \frac{\delta q V_1}{2} - \frac{\delta q V_2}{2} = \frac{\delta q (V_1 - V_2)}{2}$$

Now, δq = the flux in the tube (by Gauss' theorem).

$= D\delta S$, where δS is the cross-sectional area of the tube, and

$V_1 - V_2 = \Sigma E \delta l$

where

δl = element of length along the flux tube

E = electric force at that point.

$$\therefore \qquad \delta W = \frac{\Sigma DE\delta S\delta l}{2} = \frac{\Sigma DE\delta v}{2}$$

where δv is the volume of the element. Similarly, the contributions from all other charge pairs can also be added up. No lines of force cross each other and each element (as shown above) belongs to one Faraday tube only.

$\therefore$ The total energy of the system

$$= \frac{\Sigma DE\delta v}{2}$$

$$= \iiint \frac{DE\delta v}{2} = \iiint \frac{\mathbf{D} \cdot \mathbf{E}\,\delta v}{2} \tag{3.3}$$

This equation holds good when the dielectric is non-uniform. The Faraday tube then gets refracted, but the arguments used above are still valid. Also, since $\mathbf{D} = \varepsilon\mathbf{E}$,

$$W = \frac{\varepsilon_0}{2} \iiint \varepsilon_r E^2 dv \tag{3.4}$$

Thus the energy appears to be distributed throughout the space of the field, with a density at any point being $(DE/2)$ J/m^3. It should be noted that this is *not* a proof that the energy is really stored in the field. It merely implies that it is justifiable to regard the energy to be stored there. 'The field energy' is thus the potential energy of the sources.

3.3.1 An Alternative Derivation for the Field Energy

Over the volume v considered, let there be a charge density ρ in the system. Then the energy of the system is

$$W_e = \frac{1}{2} \iiint V\rho dv \tag{3.5}$$

where V is the final potential of the element considered. Now the charge density, ρ = div $\mathbf{D}$, by Gauss' theorem.

$$\therefore \qquad W_e = \frac{1}{2} \iiint V \text{ div } \mathbf{D}\, dv \tag{3.6}$$

From vector analysis, we have

$$\text{div } (V\,\mathbf{D}) = V \text{ div } \mathbf{D} + \mathbf{D} \cdot \text{grad } V$$

$$V\rho = V \text{ div } \mathbf{D} = \text{div } (V\,\mathbf{D}) - \mathbf{D} \cdot \text{grad } V \qquad \text{and} \qquad E = -\text{ grad } V$$

$$\therefore \qquad V\rho = \text{div } (V\,\mathbf{D}) + \mathbf{D} \cdot \mathbf{E}$$

$$\therefore \qquad W_e = \frac{1}{2} \iiint_v \text{div } (V\,\mathbf{D}) dv + \frac{1}{2} \iiint_v \mathbf{D} \cdot \mathbf{E}\, dv \tag{3.7}$$

The volume v in Eqs. (3.5) and (3.7) must enclose all the charges in the system, but otherwise is quite arbitrary. This is because, when there is no charge (i.e. $\rho = 0$) the integration of Eq. (3.5) does not affect the final result. So we can assume v to be a large sphere with its centre

somewhere in the system. Hence the first integral of Eq. (3.7) can be converted into a surface integral, i.e.

$$\iiint_v \text{div}(V\,\mathbf{D})\,dv = \oiint_S V\,\mathbf{D}\cdot d\mathbf{S}$$

where S is the surface enclosing volume v. The above equation is due to Gauss' theorem.

Over large distances, V decreases as $(1/r)$ and $\mathbf{D}$ decreases as $(1/r^2)$ whereas S increases as r^2. So as $r \to \infty$, the surface integral would tend to zero, and so Eq. (3.7) becomes

$$W_e = \frac{1}{2}\iiint_v \mathbf{D}\cdot\mathbf{E}\,dv \tag{3.8}$$

This is the expression for the energy of the static electric field expressed as an integral over the whole field. Again it should be noted that this is *not* a proof that the energy is stored in the vacuum (free space). However it will be proved later that the energy is located at all points in the field (including vacuum) for time-varying fields.

3.4 FORCES ON CONDUCTORS AND DIELECTRICS

It is easier to determine the forces on charged conductors compared with those on dielectrics because in the former problems we need to consider only the free charges on the conductors whereas for the latter there are additional forces due to polarization effects. So we shall first consider the conductors.

3.4.1 Forces and Pressures on Conductors

Force and torque are both associated with energy transfer. Since W is the assembly work, the force in a direction, say, x is given by

$$F_x = -\frac{\partial W}{\partial x} \tag{3.9}$$

i.e. by the rate of change of energy in the x-direction. Note that this is not the time-rate of change.

If this force is expressed in vector form, then

$$\mathbf{F} = -\text{grad}\ W \tag{3.10}$$

Let us consider, in an electrostatic field, the electric field intensity $\mathbf{E}$ on the surface of a conductor having a surface charge density ρ_S,

$$E = E_n = \frac{\rho_S}{\varepsilon_0} \tag{3.11}$$

This is the boundary condition for E_n on the conductor surface as obtained in Eq. (2.62). From this, we then find the pressure on the conductor surface. In this system, the total electric field $\mathbf{E}$ both inside and outside the conductor is the sum of the $\mathbf{E}$ fields created by the surface charges in the immediate vicinity of the point under consideration ($= \mathbf{E}_{\rho_S}$) and all other charges in the system ($= \mathbf{E}_S$). The first component $\mathbf{E}_{\rho_S}$ is perpendicular to the surface but away from it from both the sides (for $+\rho_S$), because for the two points M and M′ close to the surface as shown in Figure 3.2, the surface charge is equivalent to a plane charged sheet. And $\mathbf{E}_S$ is obviously the same for both M and M′.

$\therefore$ At M′ $\quad -\mathbf{E}_{\rho_S} + \mathbf{E}_S = 0$

and at M, $\quad \mathbf{E}_{\rho_S} + \mathbf{E}_S = \mathbf{E}$

$\therefore \quad \mathbf{E}_S = \mathbf{E}_{\rho_S} = \dfrac{\mathbf{E}}{2}$ (3.12)

Consider now a cylindrical Gaussian surface S (on the interface under consideration), on the two ends of which are located the points M and M′ (Figure 3.3). Take (div **E**) over the closed surface S of cross-sectional area A.

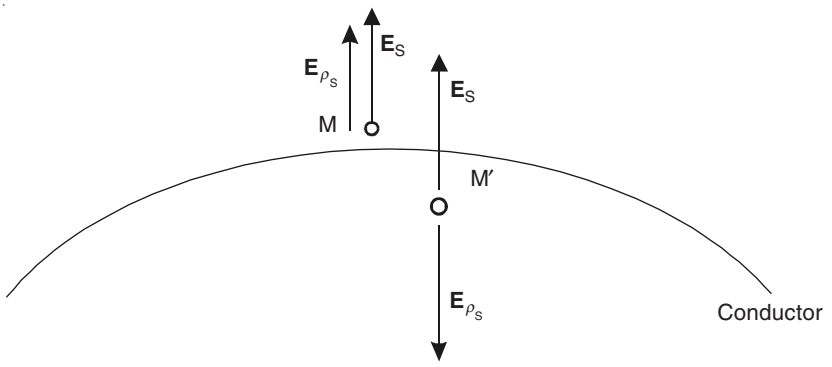


Figure 3.2 Components of *E* field near the surface of a charged conductor.

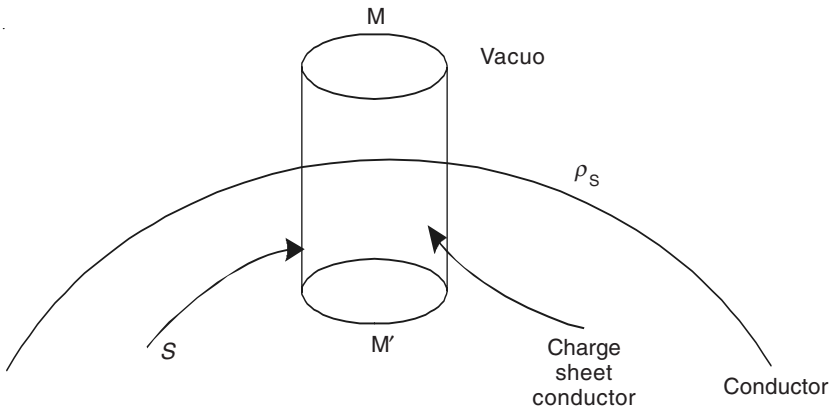


Figure 3.3 Gaussian surface across the charged conductor surface.

$\therefore \quad (E_{\rho_S}$ at M + E_{ρ_S} at M′)$A = \dfrac{\rho_S}{\varepsilon_0} A$, both being outwards

$\mathbf{E}_S$ being the same at both M and M′.

$\therefore \quad \mathbf{E}_{\rho_S}$ at M = $-\mathbf{E}_{\rho_S}$ at M′ = $\dfrac{\rho_S}{2\varepsilon_0}$, in magnitude only (3.13)

$\therefore$ Pressure on the conductor surface

$$p = \frac{dF}{dS} = \rho_S E_S = \rho_S \left(\frac{\rho_S}{2\varepsilon_0} \right) = \frac{\rho_S^2}{2\varepsilon_0} \tag{3.14}$$

Since $\rho_S = D = \varepsilon_0 E$ on the surface of the charged conductor,

$$\therefore \quad p = \frac{\varepsilon_0 E^2}{2} = \frac{\mathbf{E} \cdot \mathbf{D}}{2} = \frac{D^2}{2\varepsilon_0} \tag{3.15}$$

$\therefore$ The total force on the charged conductors (situated in vacuum) is

$$\mathbf{F} = \oiint_S p \, d\mathbf{S} = \frac{1}{2} \oiint_S (\mathbf{E} \cdot \mathbf{D}) \, d\mathbf{S} \tag{3.16}$$

These equations can be used to evaluate the force acting on one plate of a parallel plate capacitor with a plate of surface area S and a distance d between them. Let the charges on the plates be $+Q$ and $-Q$. The force F on the plate can be calculated by either method. The pressure on the inner surface of each plate (neglecting the fringing effect) is

$$p = \frac{\varepsilon_0 E^2}{2} = \frac{\rho_S^2}{2\varepsilon_0} = \frac{Q^2}{2S^2\varepsilon_0} \tag{3.17}$$

[(refer to Eq. (1.53)]

On the outer surface, $p = 0$.

$$\therefore \quad \mathbf{F} = \left(\frac{Q^2}{2S^2\varepsilon_0} \right) \mathbf{S} = \frac{Q^2}{2S\varepsilon_0} \mathbf{u}_n \tag{3.18}$$

where $\mathbf{u}_n$ is the unit normal vector to the inner surface of the plate. Hence the force is attractive.

3.5 ELECTROSTATIC FORCES ON DIELECTRICS

SOME GENERAL COMMENTS

When dielectric materials are present in the electric field, in addition to the conducting bodies, evaluating the forces in the system becomes very complex. This is because we have to determine the location of all the charges in the system and also the polarization vector, even if there is no net polarization effect. Also, most of the real dielectrics are compressible. Hence the permittivity and the polarization charges are functions of the stresses of the dielectrics. This phenomenon is called 'electrostriction'. This is a vicious circle because the volume forces in a dielectric are functions of the polarization charges and the polarization charges are, to some extent, the functions of the dielectric stresses.

However it is possible to determine the total forces and the torques on the separate bodies of the system by the method of virtual displacement. Also since the electrostatic forces are very small, their applications are restricted mostly to delicate electrical instruments, and it is mostly necessary to have the knowledge of the total forces in the system.

3.6 GENERAL METHOD OF DETERMINING FORCES IN ELECTROSTATIC FIELDS

When the total forces and the torques on the bodies in electrostatic fields are required, the method used is based on the 'principle of virtual displacement' which in turn is based on 'the law of conservation of energy'. This implies that, in electrostatic fields, the electric forces on the bodies

are balanced by equal mechanical forces, and hence are stationary. This is because, by definition, all charged bodies and dielectrics in electrostatic fields are static. Let us imagine now that an electric force moved a charged or dielectric body by an infinitesimal distance δx, then this imaginary (or virtual) displacement would be accompanied by virtual work δW, such that

$$F_x = \frac{\delta W}{\delta x} \tag{3.19}$$

where F_x is the component of the total electric force **F** in the direction of the displacement δx.

The virtual work δW can be determined in terms of the stored energy in the field by considering the following two special cases:

1. When the charges on all the bodies, during the virtual displacement, stay constant
2. When the potentials of all the bodies, during the virtual displacement, stay constant.

We consider these two cases separately.

1. Charges on all the conducting bodies are kept constant (i.e. conductors are isolated). Hence no energy is supplied to the system. So the virtual work done by the electric forces according to the law of conservation of energy must be accompanied by an equal amount of change of energy of the system δW_e. It is assumed that there is no loss of energy associated with the virtual displacement.

$$\therefore \qquad \delta W = -\delta W_e \tag{3.20}$$

Since no energy is supplied to the system the energy of the system was larger before the displacement, and thus the energy of the system has decreased. Hence δW is negative. Therefore the component F_x (in the direction of x) of the electric force **F** on any body of the system is

$$F_x = -\frac{\delta W_e}{\delta x} \text{ (charges kept constant)} \tag{3.21}$$

Similar argument holds for the rotating bodies. The virtual work δW done by the z-component of the total torque **T** in rotating a body through a virtual angle $\delta\phi$ around the z-axis of rotation is

$$\delta W = T_z \delta\phi \tag{3.22}$$

Combining this with Eq. (3.20), we get

$$T_z = -\frac{\delta W_e}{\delta\phi} \text{ (charges kept constant)} \tag{3.23}$$

2. The potentials of all the conducting bodies are kept constant during the virtual displacement, i.e. all the conducting bodies are connected to the sources which supply the charges during the displacement, so that all the potentials remain at the previous values in spite of the somewhat different geometry of the system after the displacement. If the potentials of the conducting bodies are V_1, V_2, ..., V_n and the supply of the charges during the virtual displacement are δQ_1, δQ_2, ..., δQ_n, then the total work δW_S done by all the sources is

$$\delta W_S = \sum_{i=1}^{n} V_i \delta Q_i \tag{3.24}$$

In this case, by the law of conservation of energy (assuming no losses), we get

$$\delta W_S = \delta W + \delta W_e \tag{3.25}$$

i.e. work done by the sources δW_S = virtual work δW + change of energy in the system.

We have from Eq. (3.1), the increase in the energy of the system if the potentials are maintained constant and the charges of the bodies changed,

$$\delta W_e = \frac{1}{2}\sum_{i=1}^{n} V_i \delta Q_i \tag{3.26}$$

$\therefore$ $\delta W_S = 2\delta W_e$ and $\delta W = \delta W_e$

$\therefore$ Exactly one-half the energy supplied by the sources is used for performing the mechanical work during the displacement. Hence the energy of the system must increase by the same amount.

$\therefore$ $$F_x = +\frac{\delta W_e}{\delta x} \text{ (potentials of conducting bodies kept constant)} \tag{3.27}$$

Similarly for the torque, we have

$$T_z = +\frac{\delta W_e}{\delta \phi} \text{ (potentials of conducting bodies kept constant)} \tag{3.28}$$

3.7 PRESSURE ON BOUNDARY SURFACES

3.7.1 Pressure on Surface of Charged Conductors

We consider a charged conductor in a homogeneous dielectric of permittivity $\varepsilon(= \varepsilon_0\varepsilon_r)$. Let the electric forces push a small surface ΔS of the conductor, for a small distance δx in a direction x, while the charge on the conductor is assumed to remain constant (see Figure 3.4).

$\therefore$ Decrease in the stored energy in the field

= the energy stored in the elemental volume $\delta v (\delta v = \Delta S \delta x \cos \alpha)$

$$= \frac{1}{2}\varepsilon E^2 \Delta S \, \delta x \cos \alpha$$

$\therefore$ $$\delta W_e = -\frac{1}{2}\varepsilon E^2 \Delta S \, \delta x \cos \alpha$$

$\therefore$ Force acting on the surface element ΔS in the direction of the x-axis is

$$\Delta F = -\frac{\delta W_e}{\delta x}$$

$$= \frac{1}{2}\varepsilon E^2 \Delta S \cos \alpha \tag{3.29}$$

$\therefore$ The component of the force parallel to the surface ($\alpha = 90°$) is zero.

$\therefore$ On the surface of the conductor, there is only the normal pressure, i.e.

$$p = \frac{\Delta F}{\Delta S} = \frac{1}{2}\varepsilon E^2 = \frac{1}{2}\mathbf{E} \cdot \mathbf{D} \tag{3.30}$$

This relation was obtained earlier, for the special case of the dielectric vacuum.

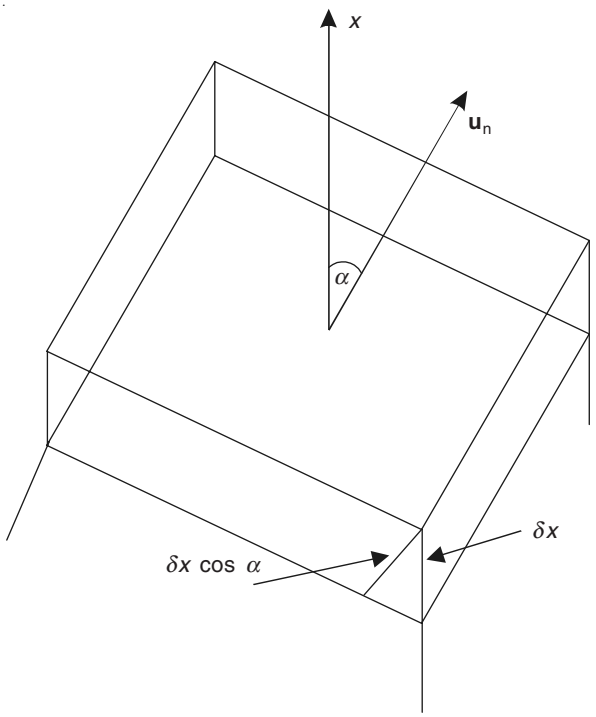


Figure 3.4 A virtual displacement for a conductor surface.

3.7.2 Pressure on Boundary Surfaces of Two Dielectrics

No tangential forces can act on the surface of two dielectrics, because a virtual displacement of the boundary parallel to itself causes no change to the electrostatic energy of the system. (This is because the system remains the same as before. So by Eqs. (3.21) and (3.27), the tangential component of the force is zero.)

The normal component of the force on the boundary surface depends on the direction of the **E** vector in both the dielectrics. So we shall now consider the effects of the two limiting directions.

1. *When the* **E** *vector is parallel to the interface* (Figure 3.5). Suppose that under the influence of the pressure, the boundary surface moved a distance δx, assuming that the potential difference between the two conductors remained unchanged.

$\therefore$ **E** will not be changed after the virtual displacement. [Refer to the boundary condition: continuity of E_t across the interface, Eq. (2.58)].

$\therefore$ Considering a cross-sectional area ΔS of the interface,

$$\delta W_e = \frac{1}{2}\varepsilon_2 E_2^2 \, \Delta S \, \delta x - \frac{1}{2}\varepsilon_1 E_1^2 \, \Delta S \, \delta x = \frac{1}{2}(\varepsilon_2 - \varepsilon_1) E_t^2 \, \Delta S \, \delta x$$

Note that at constant potential, $F_x = +\dfrac{\delta W_e}{\delta x}$.

$\therefore$
$$p_{Et} = \left(\frac{1}{\Delta S}\right)\left(\frac{\delta W_e}{\delta x}\right) = \frac{1}{2}(\varepsilon_2 - \varepsilon_1) E_t^2 \tag{3.31}$$

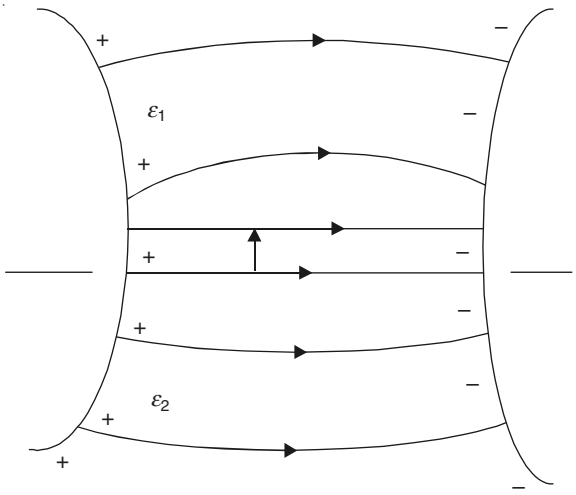


Figure 3.5 An interface between two dielectrics parallel to the **E** vector.

directed from the dielectric 2 to the dielectric 1. This is the force per unit area acting from the dielectric ε_2 to the dielectric ε_1.

The pressure tends to push the boundary surface towards the dielectric of smaller permittivity.

2. *When the* **E** *vector is perpendicular to the interface* (Figure 3.6). Let the influence of the pressure cause the boundary surface to move a small distance δx from the dielectric

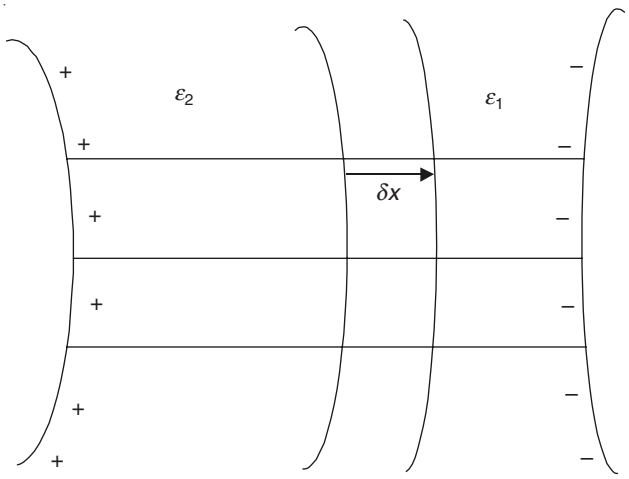


Figure 3.6 An interface between two dielectrics perpendicular to the **E** vector.

of permittivity ε_2 to the dielectric of permittivity ε_1. Due to this virtual displacement of the boundary, there will be an associated virtual change of energy, δW_e, given by

$$\delta W_e = \left(\frac{D_2^2}{2\varepsilon_2} - \frac{D_1^2}{2\varepsilon_1} \right) \Delta S \, \delta x$$

where ΔS is the surface element of the boundary. This change assumes that the free charges on the two conductors remain unchanged. The boundary condition of the interface would be $D_1 = D_2 = D_n$, since there are no charges on the interface.

$$\therefore \qquad \delta W_e = \frac{1}{2}\left(\frac{1}{\varepsilon_2} - \frac{1}{\varepsilon_1}\right) D_n^2 \, \Delta S \, \delta x$$

$$\therefore \quad \text{The pressure on the interface} = p_{D_n} = -\left(\frac{1}{\Delta S}\right)\left(\frac{\delta W_e}{\delta x}\right)$$

$$= \frac{1}{2}\left(\frac{1}{\varepsilon_2} - \frac{1}{\varepsilon_1}\right) D_n^2 \tag{3.32}$$

directed from the dielectric 2 to the dielectric 1.

$\therefore$ The pressure acts in the direction of dielectric of smaller permittivity.

From Eqs. (3.31) and (3.32), we generalize the expressions for the pressure on the interface when it makes any angle with the **E** vector, i.e.

$$p = \frac{1}{2}(\varepsilon_2 - \varepsilon_1)\left(E_t^2 + \frac{D_n^2}{\varepsilon_1 \varepsilon_2}\right) \tag{3.33}$$

directed from the dielectric of permittivity ε_2 to the dielectric of permittivity ε_1.

As an example, consider the pressure on the surface of the liquid dielectric, when the plates of a parallel plate capacitor are partially dipped in a homogeneous fluid dielectric, as shown in (Figure 3.7). If ε is the permittivity of the liquid dielectric, then the upward pressure on the surface of the liquid dielectric will be

$$= \frac{1}{2}(\varepsilon - \varepsilon_0) E_t^2$$

$$= \frac{1}{2}(\varepsilon - \varepsilon_0)\left(\frac{V_1 - V_2}{d}\right)^2 \tag{3.34}$$

where $V_1 - V_2$ is the P.D. between the plates and d is the gap between them. It should be noted

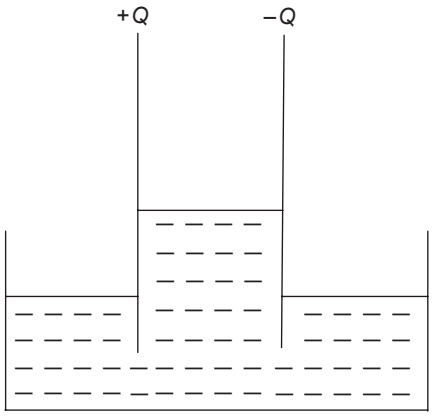


Figure 3.7 Forces pushing the liquid dielectric upwards.

that the origin of this pressure is not on the boundary surface, since the polarization charges are not present; nor is the field non-uniform. Thus whilst the pressure is evaluated correctly, these expressions say nothing about the location of the forces.

3.8 STABILITY OF THE ELECTROSTATIC SYSTEM (EARNSHAW'S THEOREM)

We have seen that the force between the electric charges obeys the inverse square law. Considering the two point charges this force can be repulsive if the charges are of the same sign, and attractive if the charges are of the opposite signs. If these two charges are regarded as a system, then we note that, for equilibrium, additional forces need to be applied to oppose the electrostatic forces which by themselves cannot produce equilibrium. This is called *Earnshaw's theorem*. The formal statement of the theorem and its proof for electrostatic systems is given below.

In 1839, Earnshaw demonstrated mathematically that, "It is impossible for a pole (or a point charge) placed in a static field of force to have a position of stable equilibrium when an inverse square law relates the force and the distance."

Proof: Consider a particle at a point (x_0, y_0, z_0) in a static field of force $\mathbf{F}(x, y, z)$.

$\therefore$ Force on the particle at this point = $\mathbf{F}(x_0, y_0, z_0)$.

If this point is to be a point of stable equilibrium, then

$$\mathbf{F}(x_0, y_0, z_0) = 0 \tag{3.35}$$

This is the condition of equilibrium; and

$$\nabla \cdot \mathbf{F}\ (x_0, y_0, z_0) < 0 \tag{3.36}$$

is the condition of stability.

If however $\mathbf{F}$ is an irrotational field, then

$$\mathbf{F}(x, y, z) = -\nabla\ \psi\ (x, y, z) \tag{3.37}$$

where ψ is a scalar potential.

It should be noted that for the electrostatic fields, the field intensity vector does satisfy the above condition with its potential V, i.e.

$$\mathbf{E} = -\nabla\ V = -\text{grad}\ V \tag{3.38}$$

$\therefore$ The necessary conditions for the stable equilibrium in terms of the scalar potential are:

$$\nabla\ \psi(x, y, z) = 0 \tag{3.39}$$

and

$$\nabla^2\ \psi(x, y, z) > 0 \tag{3.40}$$

These conditions follow from Eqs. (3.35) and (3.36). Now in a charge-free region of the electrostatic field, the field intensity vector $\mathbf{E}(x, y, z)$ is solenoidal as well as irrotational.

$\therefore$
$$\nabla \cdot \mathbf{E}(x, y, z) = 0 \tag{3.41}$$

and

$$\nabla \times \mathbf{E}(x, y, z) = 0 \tag{3.42}$$

This shows that Eq. (3.38) is a consequence of Eq. (3.42), and combining Eqs. (3.38) with (3.41), we get

$$\nabla^2 V(x, y, z) = 0 \tag{3.43}$$

which indicates that the electrostatic potential V does not satisfy Eq. (3.40) and hence does not fulfil the condition of stability.

This is true for any charged particle of charge Q placed in the field, i.e.

$$\mathbf{F}(x, y, z) = Q\mathbf{E}(x, y, z),$$

$$\therefore \quad \nabla \cdot \mathbf{F}(x, y, z) = 0$$

$\therefore$ A charged particle cannot rest in stable equilibrium under the influence of the electric field alone.

This proof, though basically deals with electrostatic fields at this stage, is quite generally applicable to all the force fields which obey the 'inverse square law'. Subsequently in 1939, this theorem was extended to magnetic fields by Braunbeck. We shall discuss these aspects of the theorem when we deal with the magnetic fields. This theorem has acquired added importance with the new applications of electromagnetic levitation schemes to transportation systems (i.e. Maglev) for both long distance as well as short distance travel using tracked vehicles.

It is to be noted that Earnshaw's theorem deals with potential energy only. For those systems which possess other types of energy as well, such as gravitational system of sun and planets possessing both kinetic and potential energy, stable equilibrium is possible.

PROBLEMS

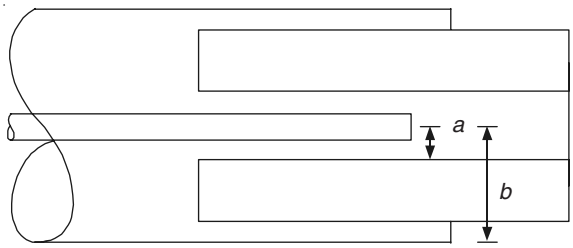
3.1 When the eight spherical drops of mercury charged to 12 V above the earth potential are coalasced into a single spherical drops (as in Problem 1.2), what is the change in the internal energy of the system?

Ans.: W_2 = Four times the original energy

3.2 Two thin metal tubes of the same length and of radii a, b, $(b > a)$ are mounted concentrically, and the inner one can slide axially within the outer one on smooth rails. Initially the inner tube is partly within the outer, when a potential difference V is applied between the tubes, it is drawn further in. Estimate the force which causes this movement, drawing attention to any assumptions required.

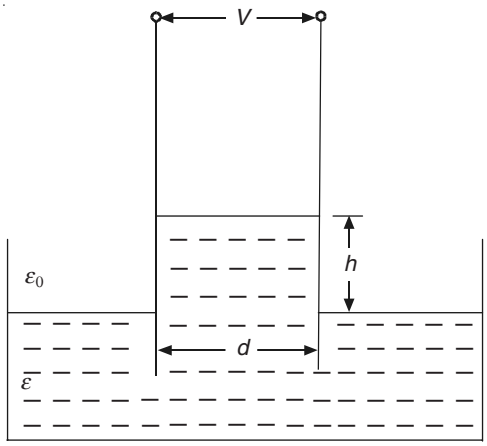
Ans.: $F = [\pi\varepsilon_0 V^2/\ln (b/a)]$ newtons

3.3 The end of a coaxial cable is closed by a dielectric piston of permittivity ε, as shown in the adjoining figure. The radii of the cable and the conductors are a and b, and the dielectric in the other part of the cable is air. What is the magnitude and the direction of the axial force acting on the dielectric piston, if the P.D. between the conductors is V?



Ans.: $F = [\pi V^2(\varepsilon - \varepsilon_0)/\ln (b/a)]$

3.4 The ends of a parallel plate capacitor are immersed vertically in a liquid dielectric of permittivity ε and mass density ρ_m. The distance between the electrodes is d and the dielectric above the liquid is air. Find h, the rise in the level of the liquid dielectric between the plates, when these are connected to a source of P.D. = V. Fringing and other side effects can be neglected



Ans.: $h = (\varepsilon - \varepsilon_0)V^2/(2d^2\rho_m g)$
where g is the gravitational constant.

4 Methods of Solving Electrostatic Field Problems

4.1 INTRODUCTION

Solving electrostatic field problems usually implies the determination of the electrostatic potential at all points in the region, from which the electric field intensity and all other quantities of interest can be obtained easily. A number of problems in which the charges are known, have been solved by using Gauss' theorem in Chapter 2. However there are other problems in which the charge distributions on some of the bodies (and not all) and the conditions on some of the boundaries might be known. In most of the electrostatic problems, it should be remembered that the potential distributions satisfy the Laplace's equation and hence these problems essentially reduce to solving this equation. There are different ways of solving this equation for different types of geometry, and we shall discuss the basic principles of some of these methods in the present chapter.

4.2 DIRECT SOLVING OF LAPLACE'S EQUATION

4.2.1 Introduction

The electrostatic potential satisfies the Laplace's equation in charge-free regions and Poisson's equation in regions containing charges. The left-hand side of both these equations is the Laplacian operator, ∇^2, operating on a scalar dependent variable, whereas the right-hand side is zero for the Laplace's equation and a nonzero quantity in the Poisson's equation. So it can be argued that the solution of the Laplace's equation is the 'complementary function' part of the general solution of the Poisson's equation $\nabla^2 V = K$, which of course is a more generalized case in which we have to evaluate the 'particular integral' part of the solution as well. However since the majority of the electrostatic problems are of Laplacian field type, we shall start our discussion with the solutions of the Laplace's equation. Furthermore, the Laplacian operator ∇^2 is the common operator of most of the equations in electromagnetism, e.g. Laplace's equation, Poisson's equation, Eddy current (or Vector diffusion type) equation, wave equation (in scalar and vector forms), general Helmholtz equation, etc. Though these are not the only equations that we find in engineering and physics, they represent a surprisingly large fraction of the equations we come

across in our study of electromagnetism and cover nearly all the major areas of present interest. All these equations are of the second order, i.e. none of them contain derivatives of higher order than the second order. Furthermore, these are all partial differential equations involving more than one independent variable and are more difficult to solve than ordinary differential equations which have only one independent variable. In general, there are two practicable methods of solution, i.e. the integral solution and the separated solution. As an example of the first method, an integral solution can be obtained for the Poisson's equation, which involves the inhomogeneous (i.e. the nonzero right-hand side of the equation) part and the distances concerned. (Such a solution is called Green's function). Solutions of similar type can also be obtained for the homogeneous equations (i.e. in which the right-hand side of the equation is zero, as in the Laplace's equation) where the integral function is determined by the boundary conditions. The integral solutions are more general in nature and are independent of the coordinate transformations, but they are not the most satisfactory type of solutions as in many cases the integral cannot be integrated in closed form and numerical values have to be obtained.

The second method of solving the linear partial differential equations (PDEs) is the method of separation of variables in which the original equation containing several independent variables is separated into a set of ordinary differential equations (ODEs), each involving one independent variable. Though this method does not have the generality of the integral solution method because of the dependence of the method on the coordinate systems, this method is more satisfactory and easy to use, where it can be used, as the solutions of the ordinary differential equations are easier to obtain than those of the partial differential equations.

It will be seen that there are an infinite number of different solutions of the Laplace's equation or that of the more general type, i.e.

$$\nabla^2\psi + k^2\psi = 0 \tag{4.1}$$

The question is not to find a solution but to find the particular solution or solutions corresponding to the particular problem we wish to solve. These various particular solutions differ from each other in the nature of the boundary conditions applied, i.e. either the geometry of the boundary varies or that the specified behaviour of the field at the boundary is different. A complete and detailed discussion of all possible types of boundaries would be very long and elaborate. So at present we shall restrict ourselves to the relevant types pertinent to the range of problems we shall be coming across in this subject.

4.2.2 Boundary Surfaces and Conditions

We have already discussed the electrostatic boundary conditions (Section 2.9) and now we shall talk about the boundary conditions in more general mathematical terms which would be applicable to all types of physical boundaries, whether they pertain to the electrostatic fields or other types, i.e. magnetostatic, electric currents, or electromagnetic. (We shall also discuss each of these types under the relevant subject heading.)

To start with there are two types of boundary surfaces—'closed' and 'open'. A closed boundary surface is one which encloses the field everywhere, restricting it to a finite volume of space. The boundary surface is outside the field and all the energy that escapes from the field is absorbed by the boundary. An open boundary is such that it does not completely enclose the field and permits it to extend to infinity in at least one direction, so that the field extends over an infinite region and with the result that the energy is absorbed by the boundary partly, the rest escaping to infinity.

The boundary conditions usually placed on the field are the specifications of the value of the field variable (which is the dependent variable of the operating equation for the field) at every point on the boundary surface, or the specification of the normal gradient of the field variable to the surface at the boundary, or a combination of both. They are:

(a) Dirichlet boundary conditions fix the value of the dependent variable, e.g. for the electrostatic field, the potential V, or ψ.
(b) Neumann boundary conditions fix the value of the normal gradient of the dependent variable, i.e. $\partial V/\partial n$ for the electrostatic potential field, or $\partial\psi/\partial n$.
(c) Cauchy conditions (or mixed boundary conditions) fix the value of both, i.e.

$$aV + b\left(\frac{\partial V}{\partial n}\right) \qquad \text{or} \qquad a\psi + b\left(\frac{\partial \psi}{\partial n}\right)$$

As we shall see later while dealing with the actual problems, each boundary condition is appropriate for different types of equations and different boundary surfaces. Physical interpretation of these boundary conditions can be made with reasonable ease while dealing with two-dimensional scalar and vector field problems, which form the largest section in electromagnetism. But when we have to deal with three-dimensional problems, and in particular with the vector variable, the interpretation of these boundaries can become quite complex, and we shall discuss them as and when we tackle such problems.

4.2.3 Coordinate Systems

We shall start our arguments based on a three-dimemsional system which is the most general practical situation and the results for the one- and two-dimensional problems are easily derivable from the three-dimensional systems as special cases of these. One way to specify the solution uniquely is, if possible, to erect on the boundary a coordinate system suitable to the boundary, i.e. we choose a coordinate system ξ_1, ξ_2, ξ_3 such that the boundary surface is one of the coordinate surfaces (say ξ_1 = a constant X). If the boundary surface is reasonably simple, it is always possible to find at least one suitable orthogonal coordinate system. If there are more than one suitable systems possible, one of them can be chosen. The partial differential equation can then be expressed in terms of the chosen coordinate system and the solutions will come out as the functions of ξ_1, ξ_2, ξ_3. These solutions will be classified according to the boundary conditions satisfied on the specified boundary surfaces. It will be found that some solutions are merely multiples of others. Such solutions should not be considered as independent solutions. There would be different solutions satisfying the surface conditions, say, $\psi = a$, $\partial\psi/\partial n = b$. These solutions can be arranged in order of the values of the ratio P, where $\psi_x(\partial y/\partial n)_x = P$. The surfaces corresponding to $P = 0$ or $\psi = 0$ are called 'nodal surfaces' (or nodes). Some of these nodal surfaces would be such that they are of the type $\xi_1 = X$, i.e. these solutions can be expressed as

$$\psi = F_1(\xi_1)\ \phi(\xi_2, \xi_3) \tag{4..2}$$

Out of the ϕ surfaces, there would be some which would be of the type

$$\psi = F_1(\xi_1)\ F_2(\xi_2)\ F_3(\xi_3) \tag{4.3}$$

Such solutions are separated into factors, each dependent on one co-ordinate variable. There are only a limited number of coordinate systems in which we can find a whole family of solutions

with their nodes behaving like this. An important property of such solutions is that all the solutions of the PDE can be built up out of the linear combinations of the members of the family of separated solutions. Those coordinate systems which allow the families of the separated solutions of a given equation, from which all solutions of the equation can be built up, are called the separable coordinate systems for the equation in question. And this method is known as the method of separation of variables.

It can be shown mathematically that there are available eleven coordinate systems in which the Laplace's equation (in three dimensions) yields three ordinary differential equations. (Solutions in one- and two-dimensions are simpler, special cases of the general three-dimensional solutions.) These solutions have to be combined so as to satisfy the boundary conditions of the problem at hand. If we succeed, then by virtue of the uniqueness theorem, the solution obtained is the required one.

Of the eleven coordinate systems mentioned above, the three most commonly known and also most widely used are the rectangular Cartesian, the cylindrical polar, and the spherical polar systems. We shall illustrate the method in connection with these three systems only and use the others with references when specific problems are explained later.

4.2.4 Separation of Variables in a Rectangular Cartesian System

The Laplace's equation in a rectangular coordinate system, according to Eq. (1.32) is

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = 0 \tag{4.4}$$

Following the method we have been discussing above, $V(x, y, z)$ can be expressed in the form

$$V(x, y, z) = X(x)\ Y(y)\ Z(z) \tag{4.5}$$

where $X(x)$ is a function of x only, $Y(y)$ is a function of y only, and $Z(z)$ a function of z only. Substituting this form of V in Eq. (4.4), we get

$$Y(y)Z(z)\frac{d^2X(x)}{dx^2} + X(x)Z(z)\frac{d^2Y(y)}{dy^2} + X(x)Y(y)\frac{d^2Z(z)}{dz^2} = 0$$

Dividing this equation by $X(x)\ Y(y)\ Z(z)$, it becomes

$$\left(\frac{1}{X(x)}\right)\frac{d^2X(x)}{dx^2} + \left(\frac{1}{Y(y)}\right)\frac{d^2Y(y)}{dy^2} + \left(\frac{1}{Z(z)}\right)\frac{d^2Z(z)}{dz^2} = 0 \tag{4.6}$$

Since each of the three terms on the left-hand side of Eq. (4.6) is a function of one coordinate variable only, and since this equation must be satisfied for all values of x, y, and z, all the three terms must equal constants. Hence Eq. (4.6) represents three ordinary differential equations, i.e.

$$\left(\frac{1}{X(x)}\right)\frac{d^2X(x)}{dx^2} = k_x^2 \tag{4.7}$$

$$\left(\frac{1}{Y(y)}\right)\frac{d^2Y(y)}{dy^2} = k_y^2 \tag{4.8}$$

$$\left(\frac{1}{Z(z)}\right)\frac{d^2Z(z)}{dz^2} = k_z^2 \tag{4.9}$$

where

$$k_x^2 + k_y^2 + k_z^2 = 0 \tag{4.10}$$

The solutions of these three equations are:

$$\begin{aligned} X(x) &= A_x \exp(k_x x) + B_x \exp(-k_x x) \quad \text{if } k_x^2 > 0 \\ X(x) &= A_x \sin k_x x + B_x \cos k_x x \quad \text{if } k_x^2 < 0 \end{aligned} \tag{4.11}$$

$$\begin{aligned} Y(y) &= A_y \exp(k_y y) + B_y \exp(-k_y y) \quad \text{if } k_y^2 > 0 \\ Y(y) &= A_y \sin k_y y + B_y \cos k_y y \quad \text{if } k_y^2 < 0 \end{aligned} \tag{4.12}$$

$$\begin{aligned} Z(z) &= A_z \exp(k_z z) + B_z \exp(-k_z z) \quad \text{if } k_z^2 > 0 \\ Z(z) &= A_z \sin k_z z + B_z \cos k_z z \quad \text{if } k_z^2 < 0 \end{aligned} \tag{4.13}$$

This can be proved by substitution. A_x, B_x, A_y, B_y, A_z, B_z are arbitrary constants of integration, which together with the constants k_x, k_y, and k_z can be determined by using the boundary conditions of the problem in hand. Usually, a single term of these functions cannot satisfy all the boundary conditions. But the sum of any number of these functions is also a solution of the Laplace's equation. So the desired solution is obtained in the form of a finite or infinite series of these terms. The process will be illustrated by some examples. For simplicity, we shall consider a two-dimensional problem first.

We wish to find the potential distribution in a rectangular strip of width a and of finite length b (Figure 4.1).

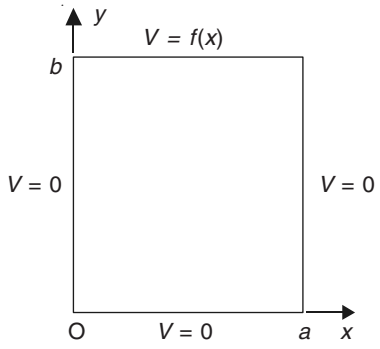


Figure 4.1 A rectangular region with known potential distributions on the boundaries.

The boundary conditions are:

$$y = b,\ V = f(x); \qquad y = 0,\ V = 0; \qquad x = a,\ V = 0; \qquad x = 0,\ V = 0 \tag{4.14}$$

Since for both the boundaries $x = 0$ and $x = a$, $V = 0$, the solutions of the form $\exp(\pm k_x x)$ from Eq. (4.11) cannot satisfy these conditions. So the solutions of $X(x)$ would be in terms of the trigonometric functions. Also, because this is a two-dimensional problem in x and y only, $k_z = 0$, and

$$k_x^2 = -k_y^2 = -k^2 \tag{4.15}$$

∴ The solution is of the form

$$V(x, y) = (A_x \sin k_x x + B_x \cos k_x x)(A_y \exp(k_y y) + B_y \exp(-k_y y)) \tag{4.16}$$

The condition $x = 0$, $V = 0$, gives $B_x = 0$; and the condition $x = a$, $V = 0$, gives $\sin ka = 0 = \sin n\pi$, where $n = 1, 2, 3, 4, \ldots$. ∴ $k = n\pi/a$.

∴ The solution reduces to the form

$$V = \sin \frac{n\pi x}{a}\left[A_y \exp\left(\frac{n\pi y}{a}\right) + B_y \exp\left(-\frac{n\pi y}{a}\right)\right] \tag{4.17}$$

Next the condition $y = 0$, $V = 0$ gives $A_y + B_y = 0$.

∴ The solution further simplifies to

$$V = A_y \sin\left(\frac{n\pi x}{a}\right)\sinh\left(\frac{n\pi y}{a}\right) \tag{4.18}$$

Since A_y is yet an arbitrary constant, it can absorb constant multipliers at each stage. Furthermore, n can take all integral values from 1 to ∞. So the general solution can be written in the form:

$$V = \sum_{n=1,2,\ldots}^{\infty} A_n \sin\left(\frac{n\pi x}{a}\right)\sinh\left(\frac{n\pi y}{a}\right) \tag{4.19}$$

Next from the boundary condition $y = b$, $V = f(x)$, we get

$$V_{y=b} = f(x) = \sum_{n=1,2,\ldots}^{\infty} A_n \sin\left(\frac{n\pi x}{a}\right)\sinh\left(\frac{n\pi b}{a}\right) \qquad \text{for } 0 < x < a \tag{4.20}$$

$$= 0 \qquad \text{for } x = 0 \text{ and } x = a$$

The coefficients A_n can be evaluated by using the concepts of Fourier series. So to evaluate the coefficients A_n for any value of n, we multiply Eq. (4.20) by $\sin(m\pi x/a)$, where m is a positive integer and then integrate the expression thus obtained, with respect to x between the limits 0 to $2a$, i.e. the period of $\sin(n\pi x/a)$. Thus,

$$\int_0^{2a} f(x) \sin\left(\frac{m\pi x}{a}\right) dx = \sum_{n=1,2,3,\ldots}^{\infty}\left[A_n \sinh\left(\frac{n\pi b}{a}\right) \times \int_0^{2a} \sin\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi x}{a}\right) dx\right] \tag{4.21}$$

Now,

$$\int_0^{2a} \sin\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi x}{a}\right) dx$$

$$= \frac{1}{2}\int_0^{2a}\left[\cos\left(\frac{(m-n)\pi x}{a}\right) - \cos\left(\frac{(m+n)\pi x}{a}\right)\right] dx$$

$$= 0 \qquad \text{for } m \neq n$$

$$= a \qquad \text{for } m = n \tag{4.22}$$

∴ From Eq. (4.21), we get

$$A_n = \frac{1}{a \sinh\left(\frac{n\pi b}{a}\right)} \int_0^{2a} f(x) \sin\left(\frac{m\pi x}{a}\right) dx \tag{4.23}$$

Substituting for A_n from Eq. (4.23) into (4.19), we get the final relation for $V(x, y)$ as

$$V = V(x, y) = \sum_{n=1,2,3,\ldots}^{\infty} \left[\frac{1}{a} \left\{ \int_0^{2a} f(x) \sin\left(\frac{m\pi x}{a}\right) dx \right\} \left\{ \frac{\sinh\left(\frac{n\pi y}{a}\right)}{\sinh\left(\frac{n\pi b}{a}\right)} \right\} \sin\left(\frac{n\pi x}{a}\right) \right] \tag{4.24}$$

For using the concept of the Fourier series, we have made the assumption that the function $f(x)$ is a repeatable pattern whose period is that of the orthogonal function sin $(n\pi x/a)$. Thus Eq. (4.24) is the unique solution of the problem for any given $f(x)$.

If $f(x) = V_0 \sin(\pi x/a)$, then $A_1 = V_0$, $A_2 = A_3 = \cdots = 0$, and the solution is

$$V = V(x, y) = V_0 \left\{ \frac{\sinh\left(\frac{\pi y}{a}\right)}{\sinh\left(\frac{\pi b}{a}\right)} \right\} \sin\left(\frac{\pi x}{a}\right) \tag{4.25}$$

If $f(x) = V_0$, then the solution is

$$V = V(x, y) = \frac{4V_0}{\pi} \sum_{n=1,3,5,\ldots}^{\infty} \left\{ \frac{\sinh\left(\frac{n\pi y}{a}\right)}{\sinh\left(\frac{n\pi b}{a}\right)} \right\} \left\{ \frac{\sin\left(\frac{n\pi x}{a}\right)}{n} \right\} \tag{4.26}$$

It should be noted that similarly the solution in three-dimensions would be a double infinite series in which not more than two coordinate variable functions can be orthogonal (i.e. in the rectangular coordinate system these would be of trigonometric type) while the remaining variable (or variables) function(s) would be of non-orthogonal type (i.e. hyperbolic). In a two-dimensional problem, the solution is a single-infinite series with one coordinate variable function being orthogonal and the other one being non-orthogonal. The choice of the orthogonal function for a particular variable is decided upon by the type of the boundary conditions of the problem. For example, if the boundary conditions on two parallel boundary surfaces normal to a particular coordinate direction, are identical, then the function of this variable (in that direction) would be orthogonal. Also, for any such problem discussed so far only one boundary can have a non zero condition (of any type). If a problem has more than one nonzero boundary, then it is solved by using the principle of superposition, i.e. only one boundary is allowed to have a nonzero condition in each sub-problem, which is then solved by the technique already explained. This is explained pictorially in Figure 4.2. Though the figure shows a problem with Dirichlet type boundaries only, the method is also equally valid for any problem with any type of boundaries, i.e. Neumann or Cauchy or any combinations of these conditions on different boundaries. These comments made so far have been with reference to the electrostatic field problems in rectangular coordinates, and yet these comments are quite general and equally valid for all Laplacian field problems in any orthogonal coordinate system, as will be seen later when we deal with practical problems.

It is to be further noted that in the series solutions which we have obtained till now, each term is a solution of the Laplace's equation and the series is required to satisfy all the boundary conditions.

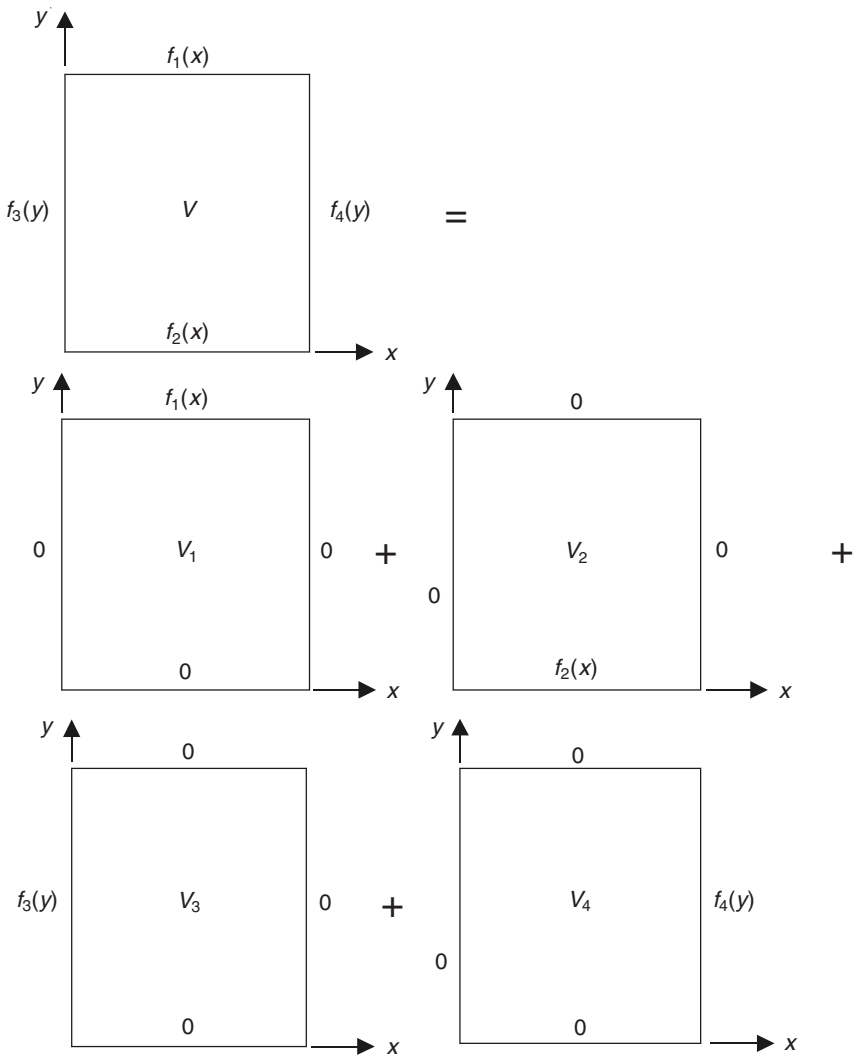


Figure 4.2 The most general case of superposition in a rectangular region.

4.2.5 Separation of Variables in a Cylindrical Polar Coordinate System

Laplace's equation in a cylindrical coordinate system has the form

$$\nabla^2 V = \left(\frac{1}{r}\right)\frac{\partial}{\partial r}\left(\frac{r\partial V}{\partial r}\right)+\left(\frac{1}{r^2}\right)\left(\frac{\partial^2 V}{\partial \phi^2}\right)+\left(\frac{\partial^2 V}{\partial z^2}\right)=0 \tag{4.27}$$

Using the procedure of separating the variables, let $V(r, \phi, z)$ be represented by the form

$$V(r, \phi, z) \equiv R(r)\,\Phi(\phi)\,Z(z) \tag{4.28}$$

Substituting from Eq. (4.28) into Eq. (4.27), and then dividing the result by $R(r)\,\Phi(\phi)\,Z(z)$, we have

$$\left(\frac{1}{rR(r)}\right)\frac{d}{dr}\left(\frac{rdR(r)}{dr}\right)+\left(\frac{1}{r^2\Phi(\phi)}\right)\left(\frac{d^2\Phi(\phi)}{d\phi^2}\right)+\left(\frac{1}{Z(z)}\right)\left(\frac{d^2Z(z)}{dz^2}\right)=0 \tag{4.29}$$

By using the arguments similar to those for Eq. (4.4), Eq. (4.29) can be satisfied for any z only if the last term on the left-hand side of this equation is a constant, i.e.

$$\left(\frac{1}{Z(z)}\right)\left(\frac{d^2Z(z)}{dz^2}\right)=k_z^2 \tag{4.30}$$

With this substitution, Eq. (4.29) can be rewritten as

$$\left(\frac{r^2}{rR(r)}\right)\frac{d}{dr}\left(\frac{rdR(r)}{dr}+k_z^2\right)+\left(\frac{1}{\Phi(\phi)}\right)\frac{d^2\Phi(\phi)}{d\phi^2}=0 \tag{4.31}$$

By similar arguments, the last term of Eq. (4.31) must be a constant, i.e.

$$\left(\frac{1}{\Phi(\phi)}\right)\left(\frac{d^2\Phi(\phi)}{d\phi^2}\right)=-k_\phi^2 \tag{4.32}$$

The negative sign of the constant is necessary because only then the function for ϕ will be orthogonal, i.e. to ensure the periodicity of the function. Then Eq. (4.31) becomes after the indicated differentiation and some rearrangement

$$\frac{d^2R(r)}{dr^2}+\left(\frac{1}{r}\right)\frac{dR(r)}{dr}+\left(k_z^2-\frac{k_\phi^2}{r^2}\right)R(r)=0 \tag{4.33}$$

This equation is known as Bessel's differential equation, and its solutions are known as the Bessel's functions of the first and the second kind respectively. Its solutions are:

$$R(r)=AJ_{k_\phi}(k_zr)+BY_{k_\phi}(k_zr) \tag{4.34}$$

where

$$J_{k_\phi}(k_zr)=\sum_{r=0}^{\infty}(-1)^r\frac{\left(\frac{k_zr}{2}\right)^{k_\phi+2r}}{\Gamma(r+1)\,\Gamma(k_\phi+r+1)} \tag{4.35}$$

where Γ is a Gamma function which is defined as

$$\Gamma(x)=\int_0^{\infty}t^{x-1}\;e^{-t}\;dt=(x-1)\,\Gamma(x-1)=(x-1)! \tag{4.36}$$

$$Y_{k_\phi}(k_zr)=\frac{\cos(k_\phi\pi)J_{k_\phi}(k_zr)-J_{-k_\phi}(k_zr)}{\sin(k_\phi\pi)} \tag{4.37}$$

For k_ϕ integer,

$$J_{-k_\phi}(k_zr)=(-1)^{k_\phi}\;J_{k_\phi}(k_zr) \tag{4.38}$$

Note: In general $J_n(x)$ and $Y_n(x)$ are the two independent solutions of the Bessel's equation

$$\frac{d^2y}{dx^2} + \left(\frac{1}{x}\right)\frac{dy}{dx} + \left(1 - \frac{n^2}{x^2}\right)y = 0 \tag{4.39}$$

$J_n(x)$ is the Bessel's function of the first kind with the argument x and of the order n; and $Y_n(x)$ is the Bessel's function of the second kind with the argument x and of the order n. For further details of the Bessel's equations and their solutions, the interested reader should refer to any book on applied maths, such as *Mathematics in Physics and Engineering* by J. Irving and N. Mullineux (Academic Press) or any other equivalent book. These Bessel's functions are orthogonal and oscillatory in nature.

So the general solution of the Laplace's equation in the cylindrical coordinate system is

$$V(r, \phi, z) = \{A J_{k_\phi}(k_z r) + B Y_{k_\phi}(k_z r)\}\{C \cos(k_\phi \phi) + D \sin(k_\phi \phi)\}\{E \cosh(k_z z) + F \sinh(k_z z)\} \tag{4.40}$$

Such terms will satisfy the Laplace equation for all values of k_ϕ and k_z and hence the most general solution will be a double infinite series with harmonic values of k_ϕ and k_z, i.e.

$$V(r, \phi, z) = \sum_{k_\phi}\sum_{k_z} \{A J_{k_\phi}(k_z r) + B Y_{k_\phi}(k_z r)\}\{C \cos(k_\phi \phi) + D \sin(k_\phi \phi)\}\{E \cosh(k_z z) + F \sinh(k_z z)\} \tag{4.41}$$

except when $k_z = 0$. For $k_z = 0$, the solution is

$$V_0 = \sum_{k_\phi} (A r^{k_\phi} + B r^{-k_\phi})\{C \cos(k_\phi \phi) + D \sin(k_\phi \phi)\}(E_0 z + F_0) \tag{4.42}$$

When both $k_\phi = 0$ and $k_z = 0$, the solution becomes

$$V_{00} = (A_{00} \ln r + B_{00})(C_0 \phi + D_0)(E_0 z + F_0) \tag{4.43}$$

Again in the solution of Eq. (4.29), the constant k_z^2 is made negative, i.e. k_z is replaced by jk_z (where $j = \sqrt{-1}$), then Eq. (4.33) becomes

$$\frac{d^2R(r)}{dr^2} + \left(\frac{1}{r}\right)\frac{dR(r)}{dr} + \left(-k_z^2 - \frac{k_\phi^2}{r^2}\right)R(r) = 0 \tag{4.44}$$

This equation is called the modified Bessel equation and its solutions are called modified Bessel functions which are denoted by $I_{k_\phi}(k_z r)$ and $K_{k_\phi}(k_z r)$ and are called the modified Bessel functions of the first kind and the second kind respectively, and are given by

$$I_{k_\phi}(k_z r) = \sum_{r=0}^{\infty} \frac{\left(\frac{k_z r}{2}\right)^{k_\phi + 2r}}{\Gamma(r+1)\,\Gamma(k_\phi + r + 1)} = j^{-k_\phi} J_{k_\phi}(jk_z r) \tag{4.45}$$

and

$$K_{k_\phi}(k_z r) = \left(\frac{\pi}{2}\right)\left(\frac{I_{-k_\phi}(k_z r) - I_{k_\phi}(k_z r)}{\sin k_\phi \pi}\right) \tag{4.46}$$

For k_ϕ integer,

$$I_{-k_\phi}(k_z r) = j^{k_\phi} J_{-k_\phi}(jk_z r) = I_{k_\phi}(k_z r) \tag{4.47}$$

Hence for the integral k_ϕ, the definition $K_{k_\phi}(k_z r)$ is treated as a limit, i.e.

$$K_{k_\phi}(k_z r) = \lim_{v \to k_\phi} \frac{I_{-v}(k_z r) - I_v(k_z r)}{\sin v\pi}$$

$$= \left[\frac{(-1)^{k_\phi}}{2} \left\{ \frac{\partial I_{k_\phi}(k_z r)}{\partial k_\phi} \right\} - \left\{ \frac{\partial I_{k_\phi}(k_z r)}{\partial k_\phi} \right\} \right] \tag{4.48}$$

The solution of Laplace's equation now becomes

$$V(r, \phi, z) = \sum_{k_\phi} \sum_{k_z} \{A I_{k_\phi}(k_z r) + B K_{k_\phi}(k_z r)\} \{C \cos(k_\phi \phi) + D \sin(k_\phi \phi)\} \{E \cos(k_z z) + F \sin(k_z z)\} \tag{4.49}$$

The choice of the solution of the Laplace's equation, i.e. Eq. (4.41) or (4.49) depends on the type of the boundary conditions in the specified problem. It should be noted that whilst the Bessel's functions $J_{k_\phi}(k_z r)$ and $Y_{k_\phi}(k_z r)$ are orthogonal functions and are oscillatory in nature, the modified Bessel's functions $I_{k_\phi}(k_z r)$ and $K_{k_\phi}(k_z r)$ are not orthogonal functions and hence are not oscillatory functions of r. So it will be seen that for the solution of Eq. (4.41), the orthogonal functions are of the variables r and ϕ and hence the nonzero boundary values would be on a constant z-plane; whereas for the solution given in Eq. (4.49), the orthogonal functions are of the variables ϕ and z, and hence the nonzero boundary conditions can be satisfied only on r = constant cylindrical surfaces. It can then be generalized that for all three-dimensional problems, the solutions, we obtain by this method, would contain orthogonal functions in not more than two variables only, and the choice of the coordinate variable for the non-orthogonal function would be decided by the location of the nonzero boundary conditions. Again when there is more than one nonzero boundary, then the problem has to be resolved into sub-problems, each containing only one nonzero boundary, and then use the method of superposition to combine the solutions of the sub-problems and thus obtain the complete solution of the original problem; similar to what has been done for the problem in the rectangular Cartesian system shown in Figure 4.2.

4.2.6 Potential Inside a Hollow Cylindrical Ring

We shall now consider a practical problem in the cylindrical coordinate system. Let us find out the potential at any point in the region bounded by the two cylinders of radii $r = a$ and $r = b$ ($a > b$) respectively, both of which are at zero potential, and the two normal planes: the plane $z = 0$ at zero potential and the plane $z = c$ whose potential is $V = f(r)$. The boundary conditions can then be summarized as

$$r = a,\ V = 0; \qquad r = b,\ V = 0; \qquad z = 0,\ V = 0; \qquad z = c,\ V = f(r) \tag{4.50}$$

Since there is no ϕ variation, this problem has axial symmetry, i.e. it is an axi-symmetric problem and we need consider only the coordinate variables r and z (Figure 4.3).

So the Laplace's equation for this problem simplifies to

$$\left(\frac{1}{r}\right) \frac{\partial}{\partial r} \left(\frac{r \partial V}{\partial r} \right) + \frac{\partial^2 V}{\partial z^2} = 0 \tag{4.51}$$

Its solution will be of the type:

$$V(r, z) = \{A J_0(k_z r) + B Y_0(k_z r)\} \{C \cosh(k_z z) + D \sinh(k_z z)\} \tag{4.52}$$

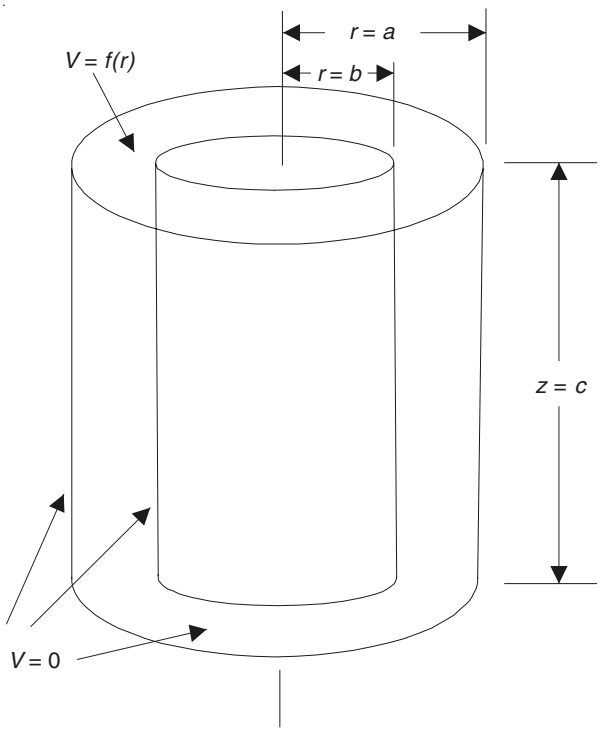


Figure 4.3 Hollow cylindrical ring with an arbitrary potential distribution $V = f(r)$ on the top surface $z = c$.

The boundary condition $z = 0$, $V = 0$ implies that the coefficients of the cosh terms must be zero. The boundary conditions $r = a$, $V = 0$ and $r = b$, $V = 0$ imply that both $r = 0$ and $r \to \infty$ are excluded and hence both $J_0(k_z r)$ and $Y_0(k_z r)$ functions are needed to satisfy these two boundaries. Hence satisfying the z = 0 boundary condition, the solution simplifies to

$$V(r, z) = \sinh (k_z z)\{AJ_0(k_z r) + BY_0(k_z r)\}$$

Next $r = b$, $V = 0$, requires the condition $AJ_0(k_z b) + BY_0(k_z b) = 0$.

$$\therefore \qquad B = -\left[\frac{AJ_0(k_z b)}{Y_0(k_z b)}\right]$$

$\therefore$ The solution further simplifies to

$$V(r, z) = A \sinh (k_z z)\left[J_0(k_z r) - \left\{\frac{J_0(k_z b)}{Y_0(k_z b)}\right\} Y_0(k_z r)\right] \tag{4.53}$$

Now to satisfy the boundary condition $r = a$, $V = 0$, the required condition is

$$J_0(k_z a) - \left\{\frac{J_0(k_z b)}{Y_0(k_z b)}\right\} Y_0(k_z a) = 0 \tag{4.54}$$

The roots of this equation will give the values of k_z, which let us denote by μ_k for the values of k.

Hence the solution is

$$V(r, z) = \sum_k A_k \sinh(\mu_k z)\left[J_0(\mu_k r) - \left\{\frac{J_0(\mu_k b)}{Y_0(\mu_k b)}\right\} Y_0(\mu_k r)\right] \tag{4.55}$$

The final boundary condition $z = c$, $V = f(r)$ will be satisfied if A_k is chosen such that

$$f(r) = \sum_k A_k \sinh(\mu_k c)\left[J_0(\mu_k r) - \left\{\frac{J_0(\mu_k b)}{Y_0(\mu_k b)}\right\} Y_0(\mu_k r)\right] \tag{4.56}$$

The term inside the square brackets is obviously a solution of the Bessel's equation. Multiply the above equation by the bracketed term, which we denote by $R_0(\mu_k r)$, for a general value of μ_k, say μ_s, then further multiply by r, and integrate the whole expression with respect to r within the limits $r = b$ to $r = a$. Since $R_0(\mu_k a) = R_0(\mu_k b) = 0$, all the terms on the right-hand side of this expression will vanish except when $\mu_k = \mu_s$. (Note the properties of the orthogonal functions.) For this term, we get

$$\int_b^R rf(r)\, R_0(\mu_k r)\, dr = A_k \left[a^2 \left(\frac{dR_0}{d(\mu_k r)}\right)^2_{r=a} - b^2 \left(\frac{dR_0}{d(\mu_k r)}\right)^2_{r=b}\right]\left[\frac{\sinh(\mu_k c)}{2}\right]$$

Differentiating $R_0(\mu_k r)$ with respect to $(\mu_k r)$ (i.e. by the argument of the function), we get

$$R_0'(\mu_k r) = -\left[J_1(\mu_k r) - \left\{\frac{J_0(\mu_k b)}{Y_0(\mu_k b)}\right\} Y_1(\mu_k r)\right] \tag{4.57}$$

$\therefore$ Solving for A_k, we get

$$A_k = 2\,\frac{\left[\int_b^a rf(r)\, R_0(\mu_k r)\, dr\right]}{\left[a^2\left\{R_0'(\mu_k a)\right\}^2 - b^2\left\{R_0'(\mu_k b)\right\}^2\right]\sinh(\mu_k c)} \tag{4.58}$$

By specifying $f(r)$ for different types of potential distributions on this surface, the distributions inside the whole annulus can be obtained.

4.2.7 Separation of Variables in a Spherical Polar Coordinate System

The three-dimensional Laplace's equation in the spherical polar coordinate system has the form

$$\frac{\partial}{\partial r}\left(\frac{r^2 \partial V}{\partial r}\right) + \left(\frac{1}{\sin\theta}\right)\frac{\partial}{\partial \theta}\left(\frac{\sin\theta\, \partial V}{\partial \theta}\right) + \left(\frac{1}{\sin^2\theta}\right)\left(\frac{\partial^2 V}{\partial \phi^2}\right) = 0 \tag{4.59}$$

Let us assume the solution to be of the form

$$V(r, \theta, \phi) = R(r)\Theta(\theta)\Phi(\phi) \tag{4.60}$$

On substituting and dividing the result by the right-hand side of the above equation, we get

$$\left(\frac{1}{R(r)}\right)\frac{d}{dr}\left(\frac{r^2 dR(r)}{dr}\right) + \left(\frac{1}{\sin\theta\Theta(\theta)}\right)\frac{d}{d\theta}\left(\frac{\sin\theta\, d\Theta(\theta)}{d\theta}\right) + \left(\frac{1}{\sin^2\theta\Phi(\phi)}\right)\left(\frac{d^2\Phi(\phi)}{d\phi^2}\right) = 0 \tag{4.61}$$

If this equation is to be satisfied for all values of r, then the first term on the left-hand side of the above equation must be a constant, i.e.

$$\left(\frac{1}{R(r)}\right)\frac{d}{dr}\left(\frac{r^2 dR(r)}{dr}\right) = k_r \tag{4.62}$$

On substituting in Eq. (4.61) and multiplying it by $\sin^2\theta$, we get

$$\left[k_r \sin^2\theta + \left(\frac{\sin\theta}{\Theta(\theta)}\right)\frac{d}{d\theta}\left(\frac{\sin\theta\, d\Theta(\theta)}{d\theta}\right)\right] + \left(\frac{1}{\Phi(\phi)}\right)\left(\frac{d^2\Phi(\phi)}{d\phi^2}\right) = 0 \tag{4.63}$$

The first term (which is inside the square brackets) is a function of θ, and the second term is a function of ϕ only. If this equation is to be satisfied for all values of θ and ϕ, each term must be equal to a constant. Since the variation of the function with respect to the coordinate variable ϕ is periodic, i.e. $V(r, \theta, \phi) = V(r, \theta, \phi + 2n\pi)$, where $n = \pm 1, \pm 2, \pm 3, ...$, we put

$$\left(\frac{1}{\Phi(\phi)}\right)\left(\frac{d^2\Phi(\phi)}{d\phi^2}\right) = -k_\phi^2 \tag{4.64}$$

Making this substitution in Eq. (4.63) and rearranging the terms, we get

$$(\sin\theta)\frac{d}{d\theta}\left(\frac{\sin\theta d\Theta(\theta)}{d\theta}\right) + \left(k_r \sin^2\theta - k_\phi^2\right)\Theta(\theta) = 0 \tag{4.65}$$

If now, the arbitrary constant k_r is written in the form

$$k_r = n(n + 1) \tag{4.66}$$

where n is a new constant, then Eq. (4.62) becomes

$$\left(\frac{1}{R(r)}\right)\frac{d}{dr}\left(\frac{r^2 dR(r)}{dr}\right) = n(n+1) \tag{4.67}$$

The solution of this equation can then be written as

$$R(r) = A_r r^n + B_r r^{-(n+1)} \tag{4.68}$$

where A_r and B_r are constants of integration. For the sake of simplicity, replacing k_ϕ by m, we get the solution of Eq. (4.64) as

$$\Phi(\phi) = A_\phi \cos m\phi + B_\phi \sin m\phi \tag{4.69}$$

except for $m = 0$, when it becomes

$$\Phi(\phi) = A_{\phi_0}\phi + B_{\phi_0} \tag{4.70}$$

If the potential function is to be single-valued with respect to the coordinate variable ϕ, then the constant m must be an integer, i.e.

$$k_\phi = m = 0, \pm 1, \pm 2, \pm 3, \pm 4, \ldots . \tag{4.71}$$

Substituting from Eqs. (4.66) and (4.71) in Eq. (4.65), we get

$$(\sin\theta)\frac{d}{d\theta}\left(\frac{\sin\theta\, d\Theta(\theta)}{d\theta}\right) + [n(n+1)\sin^2\theta - m^2]\,\Theta(\theta) = 0 \tag{4.72}$$

Before considering a more general solution of Eq. (4.72), we consider the most important special case in which V (the potential) is independent of the variable ϕ, so that Φ is constant and $m = 0$. This equation then becomes, on substituting $\cos\theta = \mu$,

$$\frac{d}{d\mu}\left[(1-\mu^2)\frac{d\Theta}{d\mu}\right] + n(n+1)\Theta = 0 \tag{4.73}$$

Equation (4.73) is called the Legendre differential equation and its solutions are called Legendre functions (or surface zonal harmonics). The two solutions are $P_n(\mu)$ and $Q_n(\mu)$, the Legendre functions (or polynomials) of the first and the second kind, respectively. It can be shown that they are of the form

$$P_n(\mu) = \sum_{s=0}^{p}(-1)^s\left[\frac{(2n-2s)!}{2^n(s!)\,(n-s)!\,(n-2s)!}\right]\mu^{n-2s} \tag{4.74}$$

where $p = n/2$ or $(n-1)/2$ whichever is an integer; and

$$Q_n(\mu) = 2^n\sum_{s=0}^{\infty}\left[\frac{(n+s)!\,(n+2s)!}{s!\,(2n+2s+1)!}\right]\mu^{-n-2s-1} \tag{4.75}$$

The details of these functions can be found in any advanced book on applied mathematics, or in *Static and Dynamic Electricity* by W.R. Smythe, or *Field Theory for Engineers* by P. Moon & D.E. Spencer, or *The Theory of Spherical and Elliptical Harmonics* by E.W. Hobson.

Both P and Q have first-order singularities at $\mu = \pm 1$, but the origin is an ordinary point, and so a series solution about the origin would be convergent within a circle of unit radius.

So far we have discussed the Laplace's equation in the spherical coordinates with V independent of ϕ. The general case of $V = V(r, \theta, \phi)$ can be handled in a similar manner. We now consider the equation (4.72) which after the substitution $\cos\theta = \mu$ becomes

$$(1-\mu^2)\frac{d^2\Theta}{d\mu^2} - 2(m+1)\frac{d\Theta}{d\mu} + (n-m)(n+m+1)\Theta = 0 \tag{4.76}$$

Its solutions are $P_n^m(\mu)$ and $Q_n^m(\mu)$ and are called the Legendre associated functions. For details, refer to the texts mentioned above. We shall now consider a problem in the spherical coordinate system.

4.2.8 Electric Field within a Charged Hollow Sphere

Let us consider the electrostatic field within a uniform sphere on the surface of which there is an arbitrary potential distribution $V = f(\theta)$. We use the spherical coordinate system with the origin at the centre of the sphere. The potential inside the sphere depends on the two coordinate variables r and θ only, and so the Laplace's equation in this case is

$$\nabla^2 V = \frac{\partial^2 V}{\partial r^2} + \left(\frac{2}{r}\right)\frac{\partial V}{\partial r} + \left(\frac{1}{r^2}\right)\frac{\partial^2 V}{\partial \theta^2} + \left(\frac{\cot\theta}{r^2}\right)\frac{\partial V}{\partial \theta} = 0 \tag{4.77}$$

This is derivable from Eq. (4.59).

The boundary condition is: $r = a$, $V = f(\theta)$ where a is the radius of the sphere. So the general form of the solution for this two-dimensional problem is

$$V = (Ar^m + Br^{-m-1})[CP_m(\cos\theta) + DQ_m(\cos\theta)] \tag{4.78}$$

Since the potential must be finite at $r = 0$ and $\theta = 0$,

$$\therefore \qquad B = 0 \qquad \text{and} \qquad D = 0$$

The solution then simplifies to

$$V = Ar^m P_m(\cos\theta) \tag{4.79}$$

To be able to satisfy the boundary condition, we use the general solution

$$V = \sum_{m=0}^{\infty} A_m r^m P_m(\cos\theta) \tag{4.80}$$

When $r = a$, $V = f(\theta) = F(\cos\theta) = F(\mu)$, using the substitution $\mu = \cos\theta$, we have

$$F(\mu) = \sum_{m=0}^{\infty} A_m r^m P_m(\mu) \tag{4.81}$$

$$\therefore \qquad A_m = \left(\frac{2m+1}{2}\right) a^{-m} \int_{-1}^{+1} F(\mu) P_m(\mu)\, d\mu \tag{4.82}$$

The solution of the problem is given by Eq. (4.80) with the values of A_m being given by Eq. (4.82).

4.3 GREEN'S FUNCTION

In Section 4.2, we obtained the solution of the electrostatic field problems by fitting the solution of the given partial differential equation (PDE) to the specified boundary conditions. The technique was that of an expansion in eigenfunctions, which is a straightforward way whenever there is a coordinate system, which suited the boundaries, in which the PDE could separate. The result would come out, usually, in the form of an infinite series. The problem of this technique was regarding the convergence of this series. This is because it is essential to add up the series to be able to get a physical insight into the overall behaviour of the solution. However nowadays with the availability of high speed computers with powerful memories, this is no longer a major problem, but in the early days when such facilities were not plentiful, it was preferable to obtain a closed function solution, even if it was an integral representation. The Green's function is such a technique.

Physically what this means is that to obtain the field produced by a distributed source, the effect of each elementary portion of the source is calculated and then all are added up. If $G(\mathbf{r}/\mathbf{r}_0)$ is the field at the observer's point $\mathbf{r}$ caused by a unit point source at the source point $\mathbf{r}_0$, then the field at $\mathbf{r}$ caused by a source distribution $\rho(\mathbf{r}_0)$ is the integral of $G\rho$ over the whole range of $\mathbf{r}_0$ occupied by the source. The function G is called the 'Green's function'. The boundary conditions can also be satisfied in the same way.

Let us consider the mathematical aspect of the Green's function before looking into its physical content. We have from Section 0.6, the Green's first identity [Eq. (0.54)] as the starting point. The scalar functions of position are U and V which are continuous and finite and are differentiable twice in the volume v which is enclosed by the surface S. Hence,

$$\iiint_v (V\nabla^2 U + \nabla V \cdot \nabla U)\, dv = \oiint_S \left(V \frac{\partial U}{\partial n}\right) dS$$

or

$$\iiint_v (V \text{ div grad } U + \text{grad } V \cdot \text{grad } U)\, dv = \oiint_S \left(V \frac{\partial U}{\partial n} \right) dS$$

Suppose now V = the electrostatic potential, and $\nabla U = \mathbf{D}$, the electrostatic flux density. Then $\nabla V = -\mathbf{E}$, where $\mathbf{E}$ is the electrostatic field intensity, and $\nabla^2 U = \rho$, the charge density. The above equation then reduces to

$$\iiint_v (V\rho - \mathbf{E} \cdot \mathbf{D})\, dv = \oiint_S (V D_n)\, dS \tag{4.83}$$

which can be rewritten as

$$\left(\frac{1}{2}\right) \iiint_v (V\rho - \mathbf{E} \cdot \mathbf{D})\, dv = \left(\frac{1}{2}\right) \oiint_S (V D_n)\, dS \tag{4.84}$$

The left-hand side of Eq. (4.84) is the difference between the assembly work of the charge density ρ and the field energy $[(1/2)\, \mathbf{E} \cdot \mathbf{D}]$. This difference is equal to a surface term given by the right-hand side. The normal component of $\mathbf{D}$, i.e. D_n can be replaced by a surface charge density ρ_S such that $D_n = -\rho_S$. Equation (4.84) can then be rewritten as

$$\left(\frac{1}{2}\right) \iiint_v V\rho\, dv + \left(\frac{1}{2}\right) \oiint_S V\, \rho_S\, dS - \left(\frac{1}{2}\right) \iint_v \mathbf{E} \cdot \mathbf{D}\, dv = 0 \tag{4.85}$$

This amounts to saying that the surface S has been replaced by S' which is outside S and the surface charge ρ_S has now been included in the region surrounded by S' (Figure 4.4). Now there is

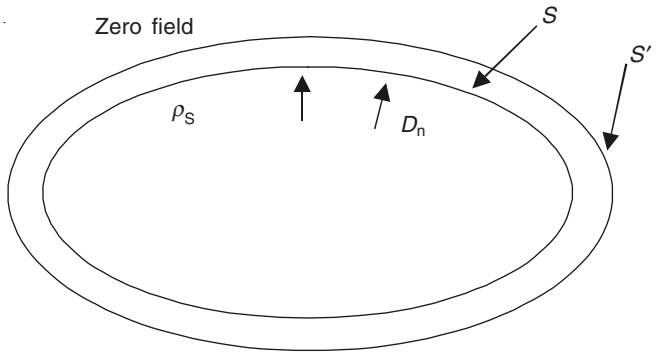


Figure 4.4 Surface charge ρ_S on the boundary S.

no contribution to the energy from the surface S' and the systems of charges ρ and ρ_S have been isolated from the rest of the space. So the internal energy of the system has been specified uniquely. Hence it can be argued that the effect of the outside world is equivalent to the surface charge $\rho_S = -D_n$.

This concept of notional surface charge ρ_S is useful as well as is required for defining the energy of the electrostatic system. If now the third term of Eq. (4.85) is transferred to the right-hand side, then we get

$$\left(\frac{1}{2}\right) \iiint_v V\rho\, dv + \left(\frac{1}{2}\right) \oiint_S V\, \rho_S\, dS = \left(\frac{1}{2}\right) \iiint_v \mathbf{E} \cdot \mathbf{D}\, dv \tag{4.86}$$

The right-hand side term of the above equation is the field energy of the system and this field energy is now found to be equal to the total assembly work and hence to the potential energy of the system. It should be noted that if the surface charge ρ_S is not included in the above equation, then neither the potential energy nor the field energy is uniquely defined for the system.

It should be further noted that there is also an alternative method of isolating the electrostatic system from the outside world. The surface potential V can be replaced by a double layer of charge which is equivalent to a step of potential. Then Eq. (4.86) can be rewritten as

$$-\left(\frac{1}{2}\right)\iiint_v \varepsilon V\,\nabla^2 V\,dv + \left(\frac{1}{2}\right)\oiint_S \varepsilon V\left(\frac{\partial V}{\partial n}\right)dS = \left(\frac{1}{2}\right)\iiint_v \mathbf{E}\cdot\mathbf{D}\,dv \tag{4.87}$$

where ε is the permittivity of the region.

So far we have considered the total energy of a system of charges. We shall now use the Green's theorem to consider the mutual energy when two or more e.s. systems are bounded by the same surface. We have the Green's theorem (or the second identity), from Eq. (0.56), as

$$\iiint_v (V\,\nabla^2 U - U\,\nabla^2 U)\,dv = \oiint_S \left(V\frac{\partial U}{\partial n} - U\frac{\partial V}{\partial n}\right)dS \tag{0.56}$$

Let the volume charge density causing the potential V be ρ and its equivalent surface charge density be ρ_S, and the corresponding quantities for the potential U be ρ' and ρ'_S, respectively, then

$$\iiint_v (V\rho' - U\rho)\,dv = -\oiint_S (V\rho'_S - U\rho_S)\,dS \tag{4.88}$$

or

$$\iiint_v V\rho'\,dv + \oiint_S V\rho'_S\,dS = \iiint_v U\rho\,dv + \oiint_S U\rho_S\,dS \tag{4.89}$$

This statement implies that the mutual energy of the two systems can be calculated either by inserting the charges ρ' and ρ'_S in the field V or by reversing the process and inserting the charges ρ and ρ_S in the field U. If the surface under consideration is at a large distance from the charges, then V and U decrease at least as $1/r$, and, $\partial V/\partial n$ and $\partial U/\partial n$ as $1/r^2$; and so the surface integral decreases as $1/r$, and therefore can be neglected. Then Eq. (4.89) simplifies to

$$\iiint_v V\rho'\,dv = \iiint_v U\rho\,dv \tag{4.90}$$

where the volume integral includes all the sources ρ and ρ'_S. In some of those problems where it may be easier to calculate one of these integrals, it might be a help in calculating the mutual capacitance of the system.

As an example, consider a small region of high charge density, the total charge being unity. This can be considered as a point unit charge which mathematically is a 'Dirac δ function', defined by the properties:

$$\delta(r) = 0,\ r \neq 0 \tag{4.91}$$

$$\int \delta(r)dv' = 1 \tag{4.92}$$

and

$$\int f(x', y', z')\,\delta(r)dv' = f(x, y, z) \tag{4.93}$$

where the point $r = 0$ is included in the volume of integration in the above two equations. The

values (x, y, z) are the coordinates of the Dirac function and f is any arbitrary function. The potential U of a small unit charge is given by

$$U = \frac{1}{4\pi\varepsilon_0 r} \tag{4.94}$$

In spherical coordinates,

$$\nabla^2 U = \left(\frac{1}{r^2}\right)\frac{\partial}{\partial r}\left(\frac{r^2 \partial u}{\partial r}\right) \tag{4.95}$$

If $r \neq 0$, then $\nabla^2 U = 0$, because there is no charge density ρ' anywhere except at $r = 0$. If $r = 0$ is included in the volume of integration (= v' enclosed by the surface S'), then

$$\iiint_{v'} \nabla^2 \left(\frac{1}{4\pi\varepsilon_0 r}\right) dv' = \frac{1}{4\pi\varepsilon_0} \oiint_{S'} \nabla\left(\frac{1}{r}\right) dS' = -\frac{1}{\varepsilon_0} \tag{4.96}$$

The charge density has the property of the Dirac function, and

$$\iiint_{v} V \rho' \, dv = V \tag{4.97}$$

i.e. the unit charge has a mutual energy with the system of charges ρ which is equal to the potential at the point where the unit charge is located. Thus the unit charge provides a means of finding the potential of a system of charges.

The potential of the unit charge, given by Eq. (4.94) is called the free space Green's function and is denoted by G.

If now we consider an unbounded region, then from the Green's theorem [Eq. (0.56)], it follows that

$$\iiint_{v} V \nabla^2 U \, dv = \iiint_{v} U \nabla^2 V \, dv \tag{4.98}$$

so that

$$V = \iiint_{v} G\rho \, dv \tag{4.99}$$

which could also have been obtained from Eq. (4.97) by transposing V and G and ρ' and ρ. These two equations are complementary, in that in Eq. (4.97) it is the energy required to insert the unit charge in the field of the charges ρ, whereas in Eq. (4.99) it is the energy required to insert ρ in the field of the unit charge. Thus when at a point of the field where the potential V is to be calculated, the Green's function provides a method of taking account of the effect of the position of the element with respect to the point under consideration.

So far we have considered an unbounded system. Now considering a bounded system, by using the Green's theorem we have

$$V = \iiint_{v} G\rho \, dv + \oiint_{S} \left\{ G\rho_S - \varepsilon_0 \left(\frac{\partial G}{\partial n}\right) V \right\} dS \tag{4.100}$$

This requires the knowledge of the potential V and its gradient $\partial V/\partial n = \rho_S/\varepsilon_0$ at the surface S, and such information may not be always available at the start of a problem. This difficulty can be partially overcome by adding to the unit charge, some surface layers of charge. Then the Green's function is modified to

$$G = \frac{1}{4\pi\varepsilon_0 r} + \chi \tag{4.101}$$

where χ is the potential of the charge layers, and within the volume v, $\nabla^2\chi = 0$. By this means, it is possible to make either G or $\partial G/\partial n$ equal to zero on S. But even then the difficulty is not completely solved. Such points are better illustrated by considering some practical problems.

To summarize what we have discussed so far, the Green's function is thus a solution for a case which is homogeneous everywhere except the source point or region . When the source is on the boundary, the Green's function can be made to satisfy the inhomogeneous boundary conditions (nonzero boundary conditions), and when it is in the region, the function may be used to satisfy the inhomogeneous equation (i.e. the Poisson's equation).

4.3.1 Green's Function for a Two-dimensional Region

This problem is similar to that solved in Section 4.2.4. We consider as before a Laplacian potential field, i.e.

$$\nabla^2 V = 0$$

in a rectangular region (Figure 4.5) which satisfies the boundary conditions

$$x = 0,\ V = 0; \qquad x = a,\ V = 0; \qquad y = b,\ V = 0; \qquad y = 0,\ V = f(x) \tag{4.102}$$

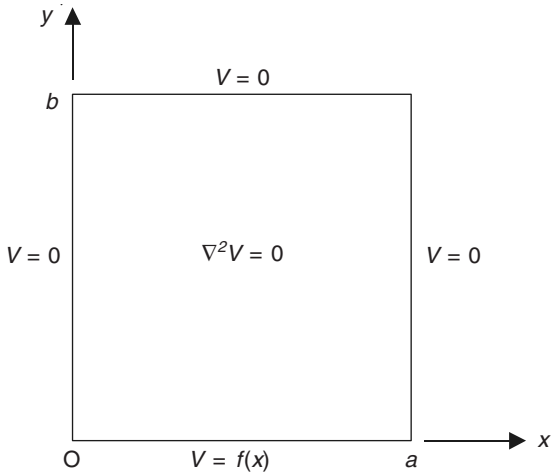


Figure 4.5 Rectangular region with specified boundaries.

Following the method of separation of variables (or by the eigenfunctions, as it is called), it will be found that the solution of this Laplacian field with the nonhomogeneous $y = 0$ boundary comes out to be the series

$$V(x, y) = \sum_{n=1,2,\ldots}^{\infty} \left[\left\{ \frac{2}{a} \int_0^a f(x) \sin\left(\frac{n\pi x}{a}\right) dx \right\} \times \frac{\sin\left(\frac{n\pi x}{a}\right) \sinh\left(\frac{n\pi(b-y)}{a}\right)}{\sinh\left(\frac{n\pi b}{a}\right)} \right] \tag{4.103}$$

Note that the Fourier coefficient integral inside the serpent brackets is *not* a function of x, but is a function of the limits a and b and hence the integrating variable x can be replaced by, say, η. This is being done so that we should be able to derive the Green's function by rearranging the solution in the integral form. Hence,

$$V(x, y) = \sum_{n=1,2,\ldots}^{\infty} \left[\left\{ \frac{2}{a} \int_0^a f(\eta) \sin\left(\frac{n\pi\eta}{a}\right) d\eta \right\} \times \frac{\sin\left(\frac{n\pi x}{a}\right) \sinh\left(\frac{n\pi(b-y)}{a}\right)}{\sinh\left(\frac{n\pi b}{a}\right)} \right] \tag{4.104}$$

and the Green's function integral solution is of the form

$$V(x, y) = \int_{S \text{ on } y=0} f(\eta)\, G\left(\frac{x, y}{\eta}\right) d\eta$$

$$= \int_0^a \left(\frac{2}{a}\right) \sum_{n=1,2,\ldots}^{\infty} f(\eta) \sin\left(\frac{n\pi\eta}{a}\right) \sin\left(\frac{n\pi x}{a}\right) \times \left[\frac{\sinh\left(\frac{n\pi(b-y)}{a}\right)}{\sinh\left(\frac{n\pi b}{a}\right)} \right] d\eta \tag{4.105}$$

Hence the Green's function will be the series

$$G\left(x, \frac{y}{\eta}\right) = \frac{2}{a} \sum_{n=1,2,\ldots}^{\infty} \frac{\sinh\left(\frac{n\pi(b-y)}{a}\right)}{\sinh\left(\frac{n\pi b}{a}\right)} \times \sin\left(\frac{n\pi\eta}{a}\right) \sin\left(\frac{n\pi x}{a}\right) \tag{4.106}$$

This is the Green's function for the boundary surface S which in this case is the boundary $y = 0$.

We can similarly write the Green's function for the boundary surface $y = b$ in the problem of Section 4.2.4 by considering Eq. (4.24) and replacing in it the variable x by η in the serpent brackets as above and then following the subsequent step, i.e.

$$G\left(x, \frac{y}{\eta}\right) = \left(\frac{2}{a}\right) \sum_{n=1,2,\ldots}^{\infty} \frac{\sinh\left(\frac{n\pi y}{a}\right)}{\sinh\left(\frac{n\pi b}{a}\right)} \times \sin\left(\frac{n\pi x}{a}\right) \sin\left(\frac{n\pi\eta}{a}\right) \tag{4.107}$$

and the Green's function solution of the problem is

$$V(x, y) = \int_{S \text{ on } y=b} f(\eta)\, G\left(\frac{x, y}{\eta}\right) d\eta \tag{4.108}$$

4.3.2 Green's Function for a Rectangular Region with Poissonian Field

In Section 4.3.1, we derived the Green's function for a homogeneous equation (Laplacian field) with a nonhomogeneous boundary. Next we consider a nonhomogeneous field, i.e. a Poissonian field with homogeneous boundaries. So let us consider the two-dimensional Poisson's equation

$$\nabla^2 V(x, y) = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = -4\pi\rho(x, y) \tag{4.109}$$

inside the rectangular boundaries (Figure 4.6):

$$x = 0,\ V = 0; \qquad x = a,\ V = 0; \qquad y = 0,\ V = 0; \qquad y = b,\ V = 0 \tag{4.110}$$

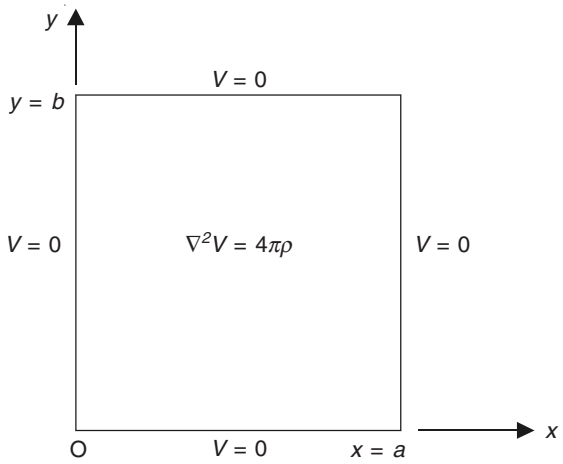


Figure 4.6 Rectangular region with Poissonian field inside.

There are two ways of solving this nonhomogeneous PDE, i.e. a single-series solution by the method described in Section 4.2.4, or by a double-infinite series solution with orthogonal functions in both x and y variables. We shall consider this solution to start with. The method is the simplest analytically though the convergence of the double series is rather slow—a computational problem which is no longer a difficult one because of the availability of today's high-speed computers. We can expand V in the terms of the double Fourier series in the form

$$V(x, y) = \sum_m \sum_n A_{mn} \sin\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right) \tag{4.111}$$

The choice of the functions of x and y are decided by the types of the boundaries specified for the problem. It is to be noted that this series is not a solution of the Laplace's equation $\nabla^2 V = 0$, but it can be shown to be a solution of the Poisson's equation (4.109). In fact, for it to be a solution of Eq. (4.109), it has to be a solution of the Laplace's equation in the specified rectangular region wherever $\rho(x, y)$ happens to be equal to zero. For example, if $\rho(x, y)$ is a delta function at the point, i.e.

$$\rho(x, y) = \delta(x - x_0)\ \delta(y - y_0) \tag{4.112}$$

then Eq. (4.111) will be a solution of the Laplace's equation at all the points in the rectangular region except for the one point, at (x_0, y_0). Coming back to the present problem, to solve Eq. (4.109), we have to expand $\rho(x, y)$ in the double Fourier series with the same periodicities, i.e.

$$\rho(x, y) = \sum_{m=1,2,\ldots}^{\infty} \sum_{n=1,2,\ldots}^{\infty} P_{mn} \sin\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right) \tag{4.113}$$

where

$$P_{mn} = \left(\frac{4}{ab}\right) \int_0^a d\xi \int_0^b d\eta\, \rho(\xi, \eta) \sin\left(\frac{m\pi\xi}{a}\right) \sin\left(\frac{n\pi\eta}{b}\right) \tag{4.114}$$

Inserting both the expressions, i.e. from Eqs. (4.113) and (4.111), in the Poisson's equation (4.109), the coefficients A_{mn} can then be determined by multiplying the substituted equation by

sin ($i\pi x/a$) and sin ($j\pi y/b$) and by integrating both sides within the limits 0 to a and 0 to b for all values of i and j, respectively. The final solution comes out in the form

$$V(x, y) = 4\pi \sum_{m=1,2,\ldots}^{\infty} \sum_{n=1,2,\ldots}^{\infty} \left[\frac{P_{mn}}{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2} \right] \sin\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right)$$

$$= \int_0^a d\xi \int_0^b d\eta\, \rho(\xi, \eta)\, G\left(\frac{x, y}{\xi, \eta}\right) \tag{4.115}$$

where

$$G\left(\frac{x, y}{\xi, \eta}\right) = \left(\frac{16\pi}{ab}\right) \sum_{m=1,2,\ldots}^{\infty} \sum_{n=1,2,\ldots}^{\infty} \left[\frac{\sin\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right) \sin\left(\frac{m\pi \xi}{a}\right) \sin\left(\frac{n\pi \eta}{b}\right)}{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2} \right] \tag{4.116}$$

is the Green's function for the inhomogeneous equation with homogeneous boundary conditions. It is to be noted that the Green's function is symmetric with respect to the interchangeability of x, y with ξ, η. Physically what this means is that the exchange of the source and the observer does not change G. This is called the *principle of reciprocity.*

We can also solve Eq. (4.109) in terms of a single-infinite series by the method of separation of variables. To solve it that way, we set

$$V(x, y) = \sum_m F_m(y) \sin\left(\frac{m\pi x}{a}\right)$$

and

$$\rho(x, y) = \sum_m \rho_m(y) \sin\left(\frac{m\pi x}{a}\right) \tag{4.117}$$

where F_m is to be evaluated, and

$$\rho_m(y) = \frac{2}{a} \int_0^a \rho(\xi, y) \sin\left(\frac{m\pi \xi}{a}\right) d\xi \tag{4.118}$$

Substituting these in Eq. (4.109), an ordinary, nonhomogeneous equation in F_m is obtained, which is

$$\frac{d^2 F_m}{dy^2} - \left(\frac{m\pi}{a}\right)^2 F_m = -4\pi\, \rho_m(y) \tag{4.119}$$

The solution of the ODE is in terms of a non-orthogonal hyperbolic function which results in the solution of the Poisson's equation in terms of a single infinite series, and the corresponding Green's function can be obtained therefrom. It needs a bit of straightforward but tedious process of expansion of the hyperbolic function of y in terms of the Fourier series in sin ($n\pi y/b$) to show that the Green's function so obtained is identical with that derived in Eq. (4.116). The details of

working out are left to the interested reader as an exercise. (For hints, the reader may refer to Morse & Feshbach: *Methods of Theoretical Physics*, Vol. 1, Chapter 7.)

4.3.3 Green's Function for an Infinite Conducting Plane

A unit line charge is placed at the point (x_0, y_0) in front of the infinite conducting plane $x = 0$, such that the line charge is parallel to the plane, and both lie in parallel y-z planes (Figure 4.7).

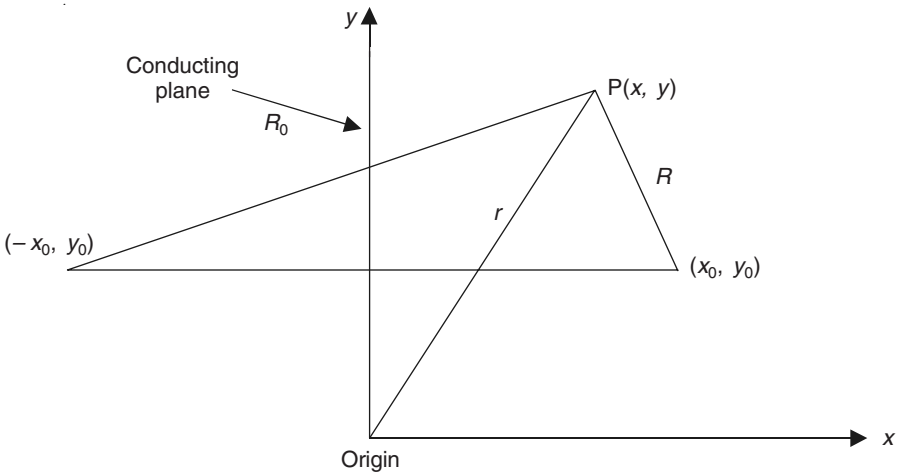


Figure 4.7 Line charge in front of a conducting plane.

The conducting metal plane must be at zero potential, and by the method of images, this is equivalent to putting a line charge of equal magnitude and opposite polarity at the point $(-x_0, y_0)$. Hence the potential at the point P [whose coordinates are (x, y)] would be

$$V_P = -\frac{1}{2\pi\varepsilon}\ln\left(\frac{R}{R_0}\right) \qquad \text{[Ref. Eq. (2.40)]} \tag{4.120}$$

The Green's function has to satisfy the homogeneous Dirichlet boundary condition for the conducting plane, i.e. for $x = 0$, $V = 0$.

$$\therefore \qquad G\left(\frac{r}{r_0}\right) = -\left(\frac{1}{2}\right)\ln\left(\frac{R}{R_0}\right), \quad x \geq 0 \tag{4.121}$$

in the positive-half of the coordinate plane. On the equipotential plane, $R = R_0$, and hence $G = 0$, i.e. the boundary condition is satisfied.

If now, instead of the line charges in front of the conducting plane we have a point charge at the point (x_0, y_0, z_0), as in Figure 4.8, then the conducting metal plane $x = 0$ and the unit positive charge at S (x_0, y_0, z_0) are again equivalent to replacing the plane by an image charge of negative nature at the point S′ $(-x_0, y_0, z_0)$—a point symmetrical to S with respect to the plane $x = 0$. Hence the potential at the observer point P (x, y, z) is

$$V_P = \frac{1}{4\pi\varepsilon}\left(\frac{1}{R} - \frac{1}{R_0}\right) \tag{4.122}$$

where R and R_0 are the distances as shown in the Figure 4.8. Once again the Green's function has to satisfy the homogeneous Dirichlet boundary condition on $x = 0$, and hence

$$G\left(\frac{r}{r_0}\right)=\left(\frac{1}{R}-\frac{1}{R_0}\right), \quad x \geq 0 \tag{4.123}$$

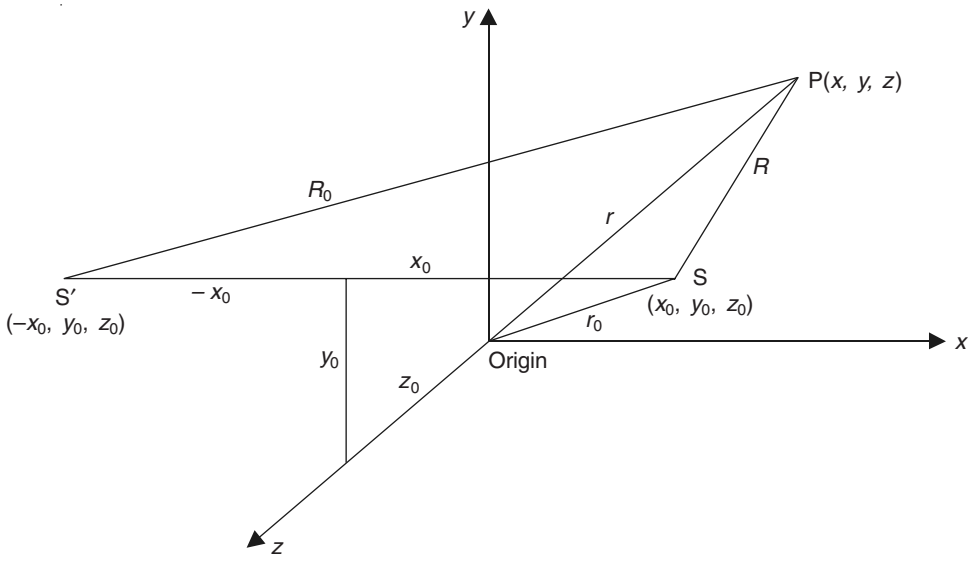


Figure 4.8 A point charge in front of an infinite conducting plane $x = 0$.

Again on the equipotential plane, $x = 0$, $R = R_0$, and hence $G = 0$. These two examples form the mathematical basis for the method of images applied to the electrostatic problems.

4.4 CONFORMAL TRANSFORMATIONS AND COMPLEX VARIABLES

Conformal transformation is one of the most powerful methods for solving the Laplacian field problems analytically and is capable of handling the boundaries of much more complicated shapes than any other analytical methods. It can be used for analyzing the fields, for instance, between non-concentric circular cables, exterior to charged conductors of polygonal sections, in slotted air-gaps of rotating machines, in waveguides of high-frequency transmission lines, and so on. The solutions are, in general, of simple forms and yield readily, the expressions for the capacitance in electrostatic fields, and the expressions for the permeance and the flux density in magnetic fields. The chief limitations to the technique are that it is applicable to two-dimensional problems only, and the boundaries have to be equipotentials or be coincident with flux lines or a combination of these two types (i.e. Dirichlet, Neumann, and mixed types).

4.4.1 Functions of Complex Variables and Conjugate Functions

Consider the complex variable

$$z = x + jy \tag{4.124}$$

where x and y are real functions and $j = \sqrt{-1}$. This quantity can be represented by a point in the complex plane with x as the abscissa and jy as the ordinate, as shown in the Figure 4.9. This z must not be confused with the z-coordinate of either the Cartesian or the cylindrical coordinate system.

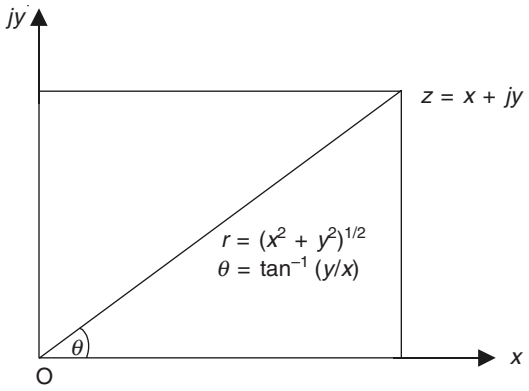


Figure 4.9 Point $z = x + jy$ in the complex plane (r is the modulus and θ the argument of the complex number z).

Then we can have the functions $W(z)$ (e.g. z^2, ln z, $1/z$, $(az + b)/(cz + d)$, and so on), and

$$W(z) = U(x, y) + jV(x, y) \tag{4.125}$$

with a real part U and an imaginary part jV, both U and V being real functions of x and y. For example,

$$W(z) = z^2 = (x + jy)^2 = (x^2 - y^2) + 2jxy \tag{4.126}$$

so that

$$U = (x^2 - y^2) \text{ and } V = 2xy \tag{4.127}$$

The function $W(z) = U + jV$ can then be represented on another complex plane with U as the abscissa and jV as the ordinate. Then we have the complex z-plane and the complex W-plane (Figure 4.10).

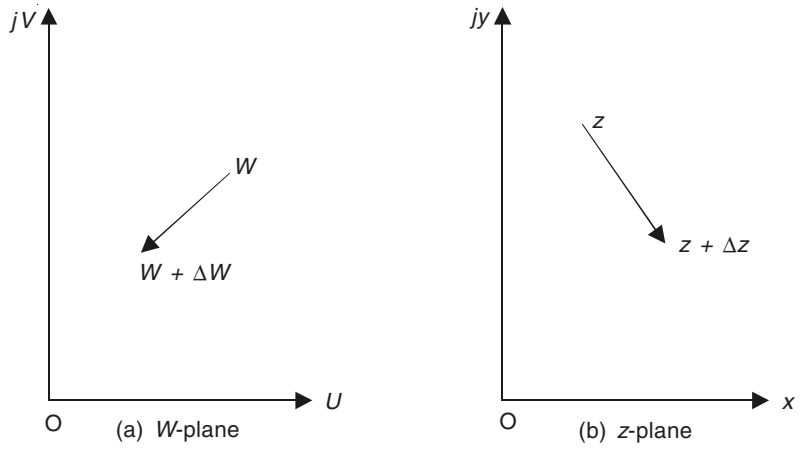


Figure 4.10 The function $W(z) = U(x, y) + V(x, y)$ in the W-plane and the derivative (dW/dz) from the changes in W and z.

Next we consider the condition necessary for the existence of the derivative (dW/dz) in the region. The point W in the W-plane corresponds to a point in the z-plane according to the specified function $W(z)$ (Figure 4.10). When z increases to $z + \Delta z$, there is a corresponding increase in W to $W + \Delta W$, where Δz and ΔW are complex quantities. The value of the derivative dW/dz is the ratio of these increments $\Delta W/\Delta z$ in the limit as $\Delta z \to 0$. Now, for a given value of z, dW/dz has to have a unique value, independent of the nature of dz. So we consider the two limiting cases, i.e. $dz = dx$ and $dz = jdy$. Hence we get

$$\frac{dW}{dz} = \frac{\partial W}{\partial x} = \frac{\partial U}{\partial x} + \frac{j\partial V}{\partial x}$$

and

$$\frac{dW}{dz} = \left(\frac{1}{j}\right)\frac{\partial W}{\partial y} = \left(\frac{1}{j}\right)\frac{\partial U}{\partial y} + \frac{\partial V}{\partial y} \tag{4.128a}$$

These two must be equal, i.e.

$$\frac{\partial U}{\partial x} = \frac{j\partial V}{\partial x} = \left(\frac{1}{j}\right)\frac{\partial U}{\partial y} + \frac{\partial V}{\partial y}$$

$$\therefore \qquad \frac{\partial U}{\partial x} = \frac{\partial V}{\partial y} \qquad \text{and} \qquad \frac{\partial U}{\partial y} = -\frac{\partial V}{\partial x} \tag{4.128b}$$

These are called the Cauchy-Riemann equations. The functions U and V related through these equations are called *conjugate functions*. A function $W(z)$ is said to be analytic, if its four partial derivatives exist and are continuous throughout the region considered, and they satisfy the Cauchy-Riemann conditions.

Note: The Cauchy-Riemann conditions have been proved for the Cartesian coordinates, but they are equally valid in other coordinate systems as well, e.g. in circular cylindrical coordinates these equations are

$$\frac{\partial U}{\partial r} = \left(\frac{1}{r}\right)\frac{\partial V}{\partial \phi} \qquad \text{and} \qquad \frac{\partial V}{\partial r} = -\left(\frac{1}{r}\right)\frac{\partial U}{\partial \phi} \tag{4.129}$$

4.4.2 Conformal Transformation

The discussion so far indicates that for the functions of the complex variables of the type

$$W(z) = f(z) \tag{4.130}$$

there is a one-to-one relationship of the points in the $W(= U + jV)$ plane to those in the $z(= x + jy)$ plane and vice versa. The W-plane can be considered to be a map or transformation or representation of the z-plane. The existence of a non-zero derivative of $W(z)$, i.e. (dW/dz) at and in the neighbourhood of z ensures that the function is analytic in the region and has unique geometrical consequences. From Figure 4.11, let W_1 and W_2 in the W-plane correspond to z_1 and z_2 in the z-plane, and also let their respective distances from W and z in their own planes be infinitesimal and the corresponding elements be dW_1, dW_2, and dz_1, dz_2. So dW_1 must be a map of dz_1, and dW_2 of dz_2; and expressing them in the polar form, we have

$$\begin{aligned} dW_1 &= dS_1 \exp(j\beta_1) & dz_1 &= ds_1 \exp(j\alpha_1) \\ dW_2 &= dS_2 \exp(j\beta_2) & dz_2 &= ds_2 \exp(j\alpha_2) \end{aligned} \tag{4.131}$$

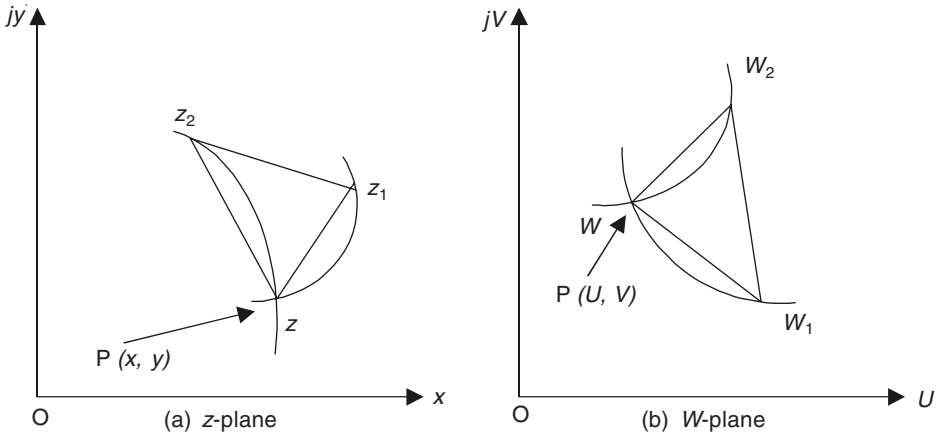


Figure 4.11 Conformal mapping.

Now, since a non-vanishing derivative dW/dz exists at this point, hence

$$\frac{dW_1}{dz_1} = \left(\frac{dS_1}{ds_1}\right) \exp\{j(\beta_1 - \alpha_1)\} = \frac{dW_2}{dz_2} = \left(\frac{dS_2}{ds_2}\right) \exp\{j(\beta_2 - \alpha_2)\} = K \exp(j\mu) \quad (4.132)$$

Since K and μ depend only on the location of the point or the value of z, hence

$$\mu = \beta_2 - \alpha_2 = \beta_1 - \alpha_1$$

$$\therefore \quad \beta_2 - \beta_1 = \alpha_2 - \alpha_1 = \alpha \text{ (say)} \quad (4.133)$$

That is, the angle between the two elements dz_1 and dz_2 is the same as that between the corresponding elements in the W-plane, dW_1 and dW_2, in both magnitude and sense. This means that the angles between the intersecting curves in the z-plane are preserved between the corresponding curves in the W-plane. In particular, when the orthogonal families of curves in the z-plane are mapped to the W-plane, they still remain orthogonal however much their finite dimensions may change. Such a transformation is called 'conformal transformation' or 'conformal mapping' or 'conformal representation'. It should be noted that any analytic function provides conformal mapping at all points where its derivative is nonzero, and the converse of this is also true.

4.4.3 Complex Potential $W(z)$

Let us evaluate the second derivative of W with respect to x and y:

$$\frac{\partial W}{\partial x} = \left(\frac{dW}{dz}\right)\left(\frac{\partial z}{\partial x}\right) = \frac{dW}{dz}; \qquad \therefore \frac{\partial^2 W}{\partial x^2} = \frac{d^2 W}{dz^2}$$

$$\frac{\partial W}{\partial y} = \left(\frac{dW}{dz}\right)\left(\frac{\partial z}{\partial y}\right) = j\frac{dW}{dz}; \qquad \therefore \frac{\partial^2 W}{\partial y^2} = -\frac{d^2 W}{dz^2}$$

$$\therefore \quad \frac{\partial^2 W}{\partial x^2} + \frac{\partial^2 W}{\partial y^2} = 0 \quad (4.134)$$

Thus the function $W(z)$ is a solution of the Laplace's equation in two dimensions. Separating the real and the imaginary parts of W, we get

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} = 0 \qquad \text{and} \qquad \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = 0 \tag{4.135}$$

Thus U and V both independently satisfy the Laplace's equation in two dimensions, and either of these two can be made to be the electrostatic potential which also satisfies the Laplace's equation in charge-free region.

If now we take the imaginary part of W, i.e. V to be the electrostatic potential, then (V = constant) will give equipotential curves. So (U = constant) curves would be orthogonal to these equipotentials, and thus they define the lines of force. The function V is then called the 'potential function', U is called the 'stream function' and W is called the 'complex potential function'.

We shall now derive the electric field intensity **E** in terms of the complex potential. In the two-dimensional field

$$E_x = -\frac{\partial V}{\partial x} \qquad \text{and} \qquad E_y = -\frac{\partial V}{\partial y}$$

and so writing in complex form, we have

$$E_x + jE_y = -\left[\frac{\partial V}{\partial x} + j\frac{\partial V}{\partial y}\right] = -\left[\frac{\partial V}{\partial x} + j\frac{\partial U}{\partial x}\right] = -\left[\frac{\partial U}{\partial y} + j\frac{\partial V}{\partial y}\right] \tag{4.136}$$

using the Cauchy-Riemann conditions. Now considering the derivative of W with respect to z, we get

$$\frac{dW}{dz} = \lim_{\Delta z \to 0} \frac{\Delta W}{\Delta z}$$

$$= \left[\frac{\left(\frac{\partial U}{\partial x} + j\frac{\partial V}{\partial x}\right)\Delta x + \left(\frac{\partial U}{\partial y} + j\frac{\partial V}{\partial y}\right)\Delta y}{\Delta x + j\Delta y}\right]\Bigg|_{\Delta x \to 0,\, \Delta y \to 0}$$

$$= E^* \text{ (i.e. complex conjugate of } E) \tag{4.137}$$

$$\therefore \qquad |E| = \left|\frac{dW}{dz}\right| \tag{4.138}$$

$\therefore$ The complex potential function completely determines a two-dimensional electrostatic field in a charge-free space.

We have seen that the stream function U is a constant along a line of force (V, being the potential function). If δs is the length element along an equipotential, then since the stream functions and the equipotentials are orthogonal, then we have

$$\frac{\partial U}{\partial s} = -\frac{\partial V}{\partial n} \tag{4.139}$$

where δn is an element of length along the line of force normal to the equipotential at the point under consideration. The negative sign above merely implies that δs is oriented to the left when viewed along δn. The above equation is a consequence of the Cauchy-Riemann conditions and becomes obvious if we visualize the x-axis parallel to δn, and the jy-axis parallel to δs.

Now, $-\dfrac{\partial V}{\partial n} = E$ (the electric field intensity at that point).

$$\therefore \qquad dU = Eds \tag{4.140}$$

The above quantity is the flux of **E** crossing the equipotential in the direction of δn through an area on the equipotential surface whose width is δs and the height is the unit of length measured on the equipotential surface perpendicular to the plane containing δn and δs. From this we can calculate the charge density on the surface of an equipotential conductor. Thus we get the flux and the potential difference from the conjugate functions $(U + jV)$.

It is being emphasized here that these quantities U, V are basically numbers having no dimensions, and we can always choose a scale constant so that the value of either V or U gives directly a value of the potential difference or a quantity of the flux. Furthermore V and U are inter-changeable, e.g. if U and V are interchanged, then Eq. (4.137) becomes $dW/dz = E$, and Eq. (4.138) still holds.

4.4.4 Some Simple Examples

We shall illustrate the power of the above method by solving a few practical problems. For the details of a large number of useful transformations, the reader is referred to standard textbooks on complex variables and applied mathematics. For a very comprehensive range of problems (both electrostatic and magnetic) the reader is recommended to refer to *Analysis and Computation of Electric and Magnetic Field Problems* by K.J. Binns & P.J. Lawrenson, *Electromagnetic Fields*, Volume 1, by E. Weber, and *A Dictionary of Conformal Representation* by H. Köber.

Since this method is useful for solving all types of Laplacian fields, it has been widely used for solving a large number of magnetic field problems. So for the present we restrict ourselves to solving a few typical electrostatic problems only.

4.4.4.1 Example 1 (Parallel plate capacitor)

Consider the field inside a parallel plate capacitor, neglecting the fringing at the ends (i.e. we assume an infinite capacitor) (see Figure 4.12.).

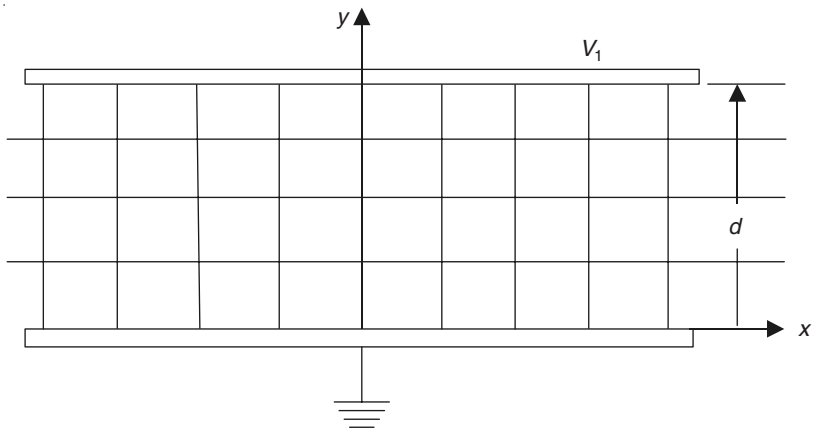


Figure 4.12 A parallel plate capacitor with its lower plate grounded and the upper plate at potential V_1. The horizontal lines are equipotentials. The vertical lines are the lines of force (stream functions). The end effects are neglected.

The equipotentials are

$$y = \text{constant, i.e. } V = \left(\frac{V_1}{d}\right)y \tag{4.141}$$

and the lines of force are

$$x = \text{constant, i.e. } U = \left(\frac{V_1}{d}\right)x \tag{4.142}$$

Hence the transformation function $W(z)$ would be

$$W(z) = U + jV = \left(\frac{V_1}{d}\right)(x + jy) = \left(\frac{V_1}{d}\right)z \tag{4.143}$$

$$\therefore \qquad E = \left|\frac{dW}{dz}\right| = \frac{V_1}{d} \tag{4.144}$$

and the charge density on the upper plate is

$$+\varepsilon_0\left(\frac{V_1}{d}\right) = +\varepsilon_0\frac{dU}{dx}$$

$$\therefore \qquad \text{Capacitance per unit area} = \frac{\text{charge density}}{V_1} = \frac{\varepsilon_0}{d} \tag{4.145}$$

In the present analysis, the fringing effects have been neglected; but in a subsequent problem we shall show how the transformation methods are capable of taking account of such effects.

4.4.4.2 *Example 2 (Two concentric cylindrical conductors)*

In spite of the above title, we shall start with a more general problem with the field between the two charged cylinders and use a two-stage transformation to show the versatility of the method. The general problem is of great value in determining the breakdown voltage of the air gap, the necessary insulation, and the capacitance of parallel cylindrical cables. Once the initial problem has been transformed to the concentric cylinders, then it becomes applicable to the coaxial cables. The two-stage transformation is from the z-plane to the t-plane and then to the W-plane. The z-t transformation is a special case of the more general bilinear transformation of the type:

$$z = \frac{at + b}{ct + d} \tag{4.146}$$

i.e.

$$z = \frac{1}{t} \qquad \text{or} \qquad t = \frac{1}{z} \tag{4.147}$$

The field between the two charged cylindrical conductors is as shown in Figure 4.13. We consider the non-concentric boundaries of this configuration, placed in the z-plane, as shown in Figure 4.14(a), with the x, y coordinates shown therein.

The origin of inversion has to be placed d units to the left of the smaller circle of radius R_2. The dimension d has to be so chosen that these two circles are transformed into concentric circles in the t-plane. The process is as follows: The transformation equation (4.147) is written as:

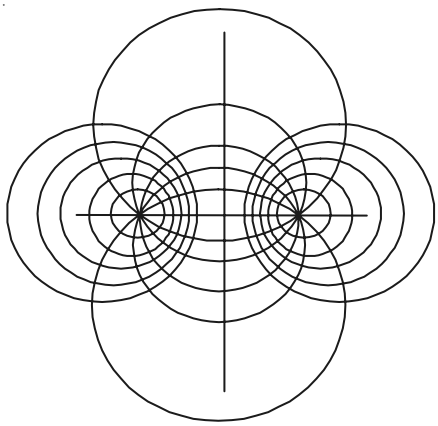


Figure 4.13 Field between two charged cylindrical conductors with uniform dielectric.

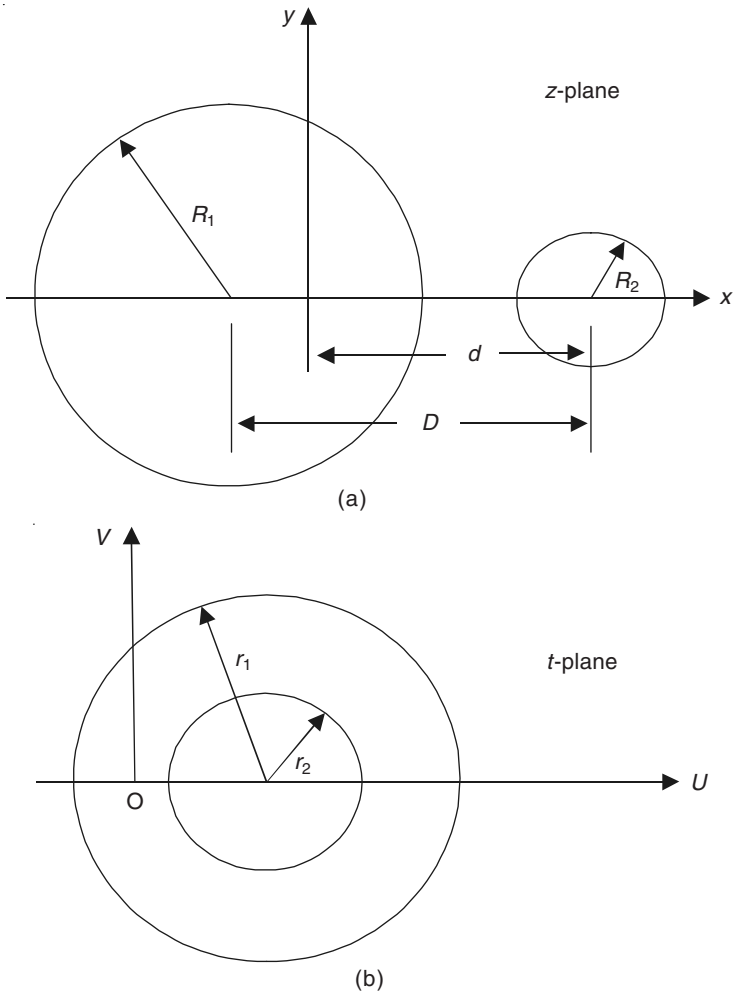


Figure 4.14 *Z*-plane to *t*-plane transformation of the two-cylinder problem.

$$t - d = \frac{1}{x + jy} - d, \text{ where } t = u + jv \tag{4.148}$$

The expression for the radius can be written as

$$R_2 = |t - d| = \left| \frac{(1 - xd) - jyd}{x + jy} \right|$$

$$\therefore \quad R_2^2 = \frac{(1 - xd)^2 + y^2 d^2}{x^2 + y^2}$$

which can be rearranged as

$$\left[x - \frac{d}{d^2 - R_2^2} \right]^2 + y^2 = \left[\frac{R_2}{d^2 - R_2^2} \right]^2 \tag{4.149}$$

i.e. the smaller circle of radius R_2 and centre $(d, 0)$ in the z-plane transforms to a circle in the t-plane, as shown in Figure 4.14(b), with its centre coordinates as $\{d/(d^2 - R_2^2), 0\}$ and radius $r_2 = R_2/(d^2 - R_2)$. Similarly the larger circle, with its centre at $(d - D, 0)$ and radius R_1 transforms itself to a circle in the t-plane with its centre at $[(d - D)/\{(d - D)^2 - R_1^2\}, 0]$, and radius $r_1 = R_1/\{(d - D)^2 - R_1^2\}$. For the two circles in the t-plane to have the same centre at $(c, 0)$, so that the circles become concentric, the required condition is

$$\frac{d}{d^2 - R_2^2} = \frac{d - D}{(d - D)^2 - R_1^2} \tag{4.150}$$

This equation fixes the required position of the origin of inversion in the z-plane by evaluating the value of d as

$$d = \frac{D^2 + R_2^2 - R_1^2}{2D} \pm \sqrt{\left(\frac{D^2 + R_2^2 - R_1^2}{2D} \right)^2 - R_2^2} \tag{4.151}$$

(Note that the value of d must be so chosen as to make r positive.)

The cylinders in the z-plane have equipotential boundaries, and in the t-plane, the concentric boundaries must also be equipotential lines with the same potential difference between them. So the field between the two concentric cylinders centred at the point $(c, 0)$, separated by a medium of relative permittivity ε_r and carrying a charge of Q units per unit length, is given by

$$W = \Psi + j\Phi = \left(\frac{Q}{2\pi\varepsilon_0\varepsilon_r} \right) \ln (t - c) \tag{4.152}$$

By considering the equipotential surfaces on the two cylinders, i.e. $t_1 = c + r_1$ and $t_2 = c + r_2$, the potential difference can be expressed as

$$\text{P.D.} = \Psi_1 - \Psi_2 = \left(\frac{Q}{2\pi\varepsilon_0\varepsilon_r} \right) \ln \left(\frac{r_1}{r_2} \right) \tag{4.153}$$

Hence the capacitance, per unit length, between the two cylinders is given by

$$C = \frac{\text{total flux per unit length } (= Q)}{\text{P.D.}} = \frac{2\pi\varepsilon_0\varepsilon_r}{\ln (r_1/r_2)} \tag{4.154}$$

To find the potential gradient, i.e. the electric field intensity **E** in the z-plane,

$$|\mathbf{E}| = \left|\frac{dW}{dz}\right| = \left|\frac{dW}{dt}\right| \times \left|\frac{dt}{dz}\right| \tag{4.155}$$

From Eq. (4.152),

$$\frac{dW}{dt} = \left(\frac{Q}{2\pi\varepsilon_0\varepsilon_r}\right)\left(\frac{1}{t-c}\right)$$

$$= \left(\frac{\Psi_1 - \Psi_2}{\ln(r_1/r_2)}\right)\left(\frac{1}{t-c}\right) \tag{4.156}$$

On substituting for Q from Eq. (4.153) and from Eq. (4.147), we have

$$\frac{dt}{dz} = -\frac{1}{z^2} \tag{4.157}$$

So Eq. (4.155) can be written as

$$|\mathbf{E}| = \left|\left(\frac{\Psi_1 - \Psi_2}{\ln(r_1/r_2)}\right)\left(\frac{1}{t-c}\right)\left(\frac{1}{z^2}\right)\right| = \left|\left(\frac{\Psi_1 - \Psi_2}{\ln(r_1/r_2)}\right)\left(\frac{1}{(1-cz)\,z}\right)\right| \tag{4.158}$$

For the consideration of the breakdown voltage between the cylinders, the maximum value of the P.D. is important. This occurs at the point on the shortest line between the cylinders at the boundary with the greater curvature. This point on the z-plane is, $z = d - R_2$ which corresponds to the point $t = c + r_2$ in the t-plane.

Hence substituting these values in Eq. (4.158), we have

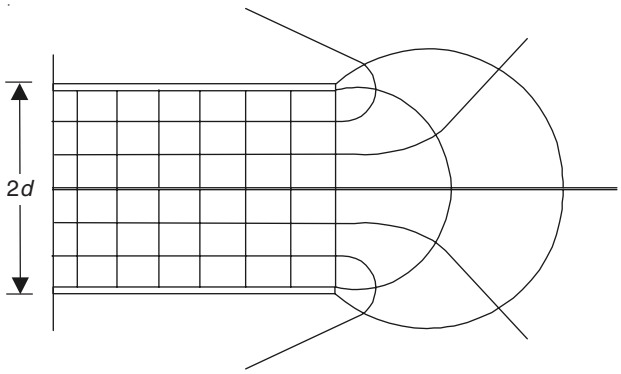
$$|\mathbf{E}|_{\max} = \left|\left(\frac{\Psi_1 - \Psi_2}{\ln(r_1/r_2)}\right)\left(\frac{1}{r_2(d-R_2)^2}\right)\right| \tag{4.159}$$

4.4.4.3 *Example 3 (A parallel plate capacitor taking account of the fringing effects at the ends)*

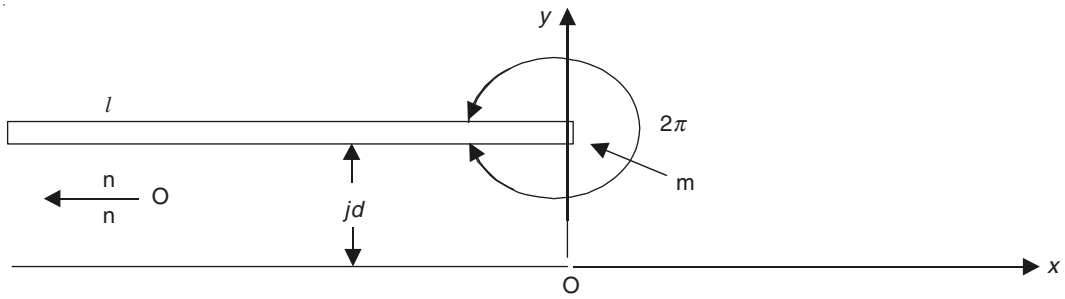
This problem is an application of the well-known Schwarz–Christoffel transformation in which the inside of a polygon is transformed into the upper-half plane (for a general detailed description of the method, the reader is recommended in particular to *The Schwarz–Christoffel Transformation and its Applications—A Simple Exposition* by Miles Walker, apart from the references previously mentioned). Hence we shall not go into the general discussion about the method, but apply it directly to this problem.

This problem is an example of transformation for a polygon with three vertices, as we shall see. The capacitor with plates of negligible thickness, which are at a distance $2d$ apart, and assumed to be charged to the potentials Ψ_1 and $-\Psi_1$, respectively [Figure 4.15(a)]. The field is symmetrical about the horizontal central line and so the top-half is represented in the z-plane as shown in Figure 4.15(b), where the real axis (x-axis) corresponds to the central line of symmetry.

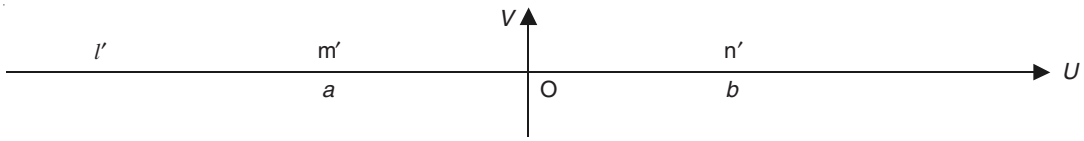
The Schwarz–Christoffel equation is used for transforming the real axis of the t-plane [Figure 4.15(c)] into the boundary of the z-plane. The upper surface of the plate, shown l m in the



(a) Fringing flux of a parallel plate capacitor



(b) Boundaries in the *z*-plane



(c) Boundaries in the *t*-plane

Figure 4.15 Transformations of the parallel plate capacitor with fringing at the ends.

z-plane is represented by the portion of the real axis of the *t*-plane between $-\infty$ and a, the lower surface of the plate (i.e. m n in the *z*-plane) by the portion between a and b. The corresponding points in the two planes then are

$$t \to -\infty,\ z \to -\infty + jd$$
$$t = a,\ z = 0 + jd$$
$$t = b,\ z \to -\infty$$
$$t \to +\infty,\ z \to +\infty + j0$$

The general equation of the Schwarz–Christoffel transformation is

$$\frac{dz}{dt} = S(t-a)^{\alpha/(\pi-1)}\,(t-b)^{\beta/(\pi-1)}\,(t-\mathrm{c})^{\gamma/(\pi-1)} \tag{4.160}$$

where $\alpha, \beta, \gamma, \delta, \ldots$ are the interior angles of the polygon at the successive vertices. The interior angles in this case are: 0 at $z \to -\infty$ and 2π at $z = 0 + jd$, and hence

$$\frac{dz}{dt} = S(t - a)(t - b)^{-1} \tag{4.161}$$

The constants S, a, and b have to be so chosen that the transformation equation, to be derived by integrating Eq. (4.161) gives the required shape and the size of the z-plane boundary. Now, since in this particular problem the z-plane configuration is independent of the specified single dimension d, any two of the three unknowns can be given convenient values and the third one being used to determine d. So to simplify Eq. (4.161), let us assume $a = -1$ and $b = 0$. Hence we get

$$\frac{dz}{dt} = \frac{S(t+1)}{t} \tag{4.162}$$

On integrating, we get

$$z = S(t + \ln t) + K \tag{4.163}$$

where S and K are constants to be evaluated. So we express Eq. (4.162) in the cylindrical polar coordinates, i.e.

$$dz = (S)\left(\frac{re^{j\theta} + 1}{re^{j\theta}}\right)\left(jre^{j\theta}\right)d\theta = jS(re^{j\theta} + 1)\,d\theta$$

Since a movement round a small circle of radius r and centre at the point $t = 0$ in the t-plane, corresponds to the movement from the real axis to the line $z = jd$ in the z-plane, hence, as r tends to zero, the above equation gives

$$\int_{-\infty}^{-\infty + jd} dz = jS \int_{0}^{\pi} d\theta \tag{4.164}$$

$$\therefore \qquad S = \frac{d}{\pi}$$

Since a has been made equal to -1, $t = 1$ corresponds to $z = jd$, and so substituting in Eq. (4.163), we get

$$jd = \left(\frac{d}{\pi}\right)\{-1 + \ln(-1)\} + K = \left(\frac{d}{\pi}\right)(-1 + j\pi) + K$$

$$\therefore \qquad K = \frac{d}{\pi}$$

and hence

$$z = \left(\frac{d}{\pi}\right)(1 + t + \ln t) \tag{4.165}$$

is the equation of transformation.

The boundaries of the z-plane have a potential difference Ψ_1 which is half of that between the capacitor plates. Hence now the upper region of the t-plane can be converted to that due to the two semi-infinite equipotential planes, one lying between 0 and ∞, and the other between 0 and $-\infty$ with a difference in potential Ψ_1 between the two halves (dividing at the origin) of the real axis. The required transformation is the logarithmic function (Figure 4.16), which is

$$W = \left(\frac{\Psi_1}{\pi}\right)\ln t \tag{4.166}$$

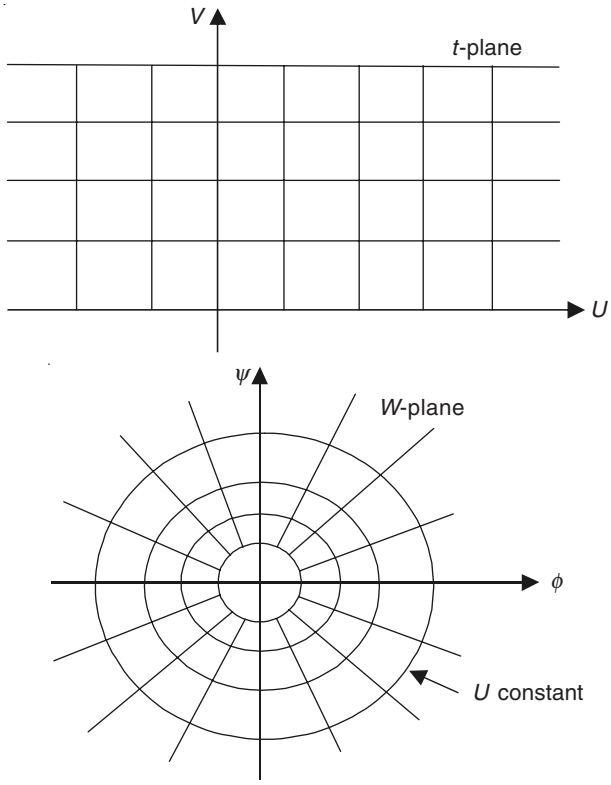


Figure 4.16 The logarithmic transformation $W = \phi + j\pi = (\Psi_1/\pi) \ln t$.

Hence the solution for the z-plane field is obtained by eliminating t between these two equations, which gives

$$z = \left(\frac{d}{\pi}\right)\left[1 + \exp\left(\frac{W\pi}{\Psi}\right) + \frac{W\pi}{\Psi_1}\right] \tag{4.167}$$

with the corresponding values as

$$z \to -\infty,\ W \to -\infty \qquad \text{and} \qquad z = jd,\ W = j\Psi_1$$

It should be noted that the choice of a and b as -1 and 0, respectively (instead of 0 and $+1$), gives the solution in its simplest form.

Capacitance. The capacitance of a parallel plate capacitor, often calculated on the assumption that the plates extend to infinity, is an underestimation. This is because of two reasons: firstly the charge density on the inner faces increases near the edges, and secondly the charge on the outer faces is neglected. Now, at any point on the surface of the plates, the charge density ρ is equal to the flux density there, i.e.

$$\rho = \left(\frac{\partial \Psi}{\partial x}\right)\varepsilon = \left|\frac{\partial W}{\partial z}\right|\varepsilon \tag{4.168}$$

where ε is the permittivity of the surrounding region.

From Eq. (4.167), by differentiation,

$$\left|\frac{dW}{dz}\right| = \left(\frac{\Psi_1}{d}\right)\left[\frac{1}{\left|\exp\left(\frac{W\pi}{\Psi}\right)+1\right|}\right]$$

Hence the charge density is

$$\rho = \frac{\rho_0}{\left|\exp\left(\frac{W\pi}{\Psi}\right)+1\right|} \tag{4.169}$$

where $\rho_0 = \Psi_1\varepsilon/d$, the charge density based on the assumption of the infinite plates. This equation shows that the charge density increases towards the edges and becomes infinite at the edge of the plate, where $z = jd$, $t = 1$, $W = j\Psi_1$.

We can also calculate the additional charge on the inner surface of the plate due to fringing, which is

$$= \int_{-\infty+jd}^{jd} (\rho - \rho_0)\, dz$$

$$= \int_{-\infty}^{j\Psi_1} \rho_0 \left[\frac{1}{1+\exp\left(\frac{W\pi}{\Psi}\right)} - 1\right]\left(\frac{d}{\Psi_1}\right)\left[\exp\left(\frac{W\pi}{\Psi}\right)+1\right] dW$$

$$= \rho_0 \frac{d}{\pi} \tag{4.170}$$

This additional charge is equivalent to a lengthening of the uniform field by the amount d/π.

The charge on the outer surface of the plates is

$$= \int_{jd}^{-\infty} \rho\, dz = \left(\frac{\rho_0 d}{\Psi_1}\right)\int_{j\Psi}^{-\infty} dW \to \infty$$

However for a capacitor of finite dimensions, it is possible to calculate this additional charge to a fair degree of accuracy by assuming the charge distribution near the edges to be identical with that near the edge of the infinite plate. This is justifiable because the charge density decreases very rapidly away from the edge, i.e. at a distance from the edge equal to $1.5d$, the value of ρ is $< \rho_0/10$. Hence

$$\text{the charge on the outer surface} = 2\int_{jd}^{-l+jd} \rho\, dz$$

where $2l$ is the width of the plate.

Note: The solution for the field of the boundary as shown in the Figure 4.15(b) was used by Rogowski to determine the electrode shapes suitable for the measurement of the breakdown strength of the gases and liquids. The electrode shape is chosen to be that of an equipotential line along which the gradient is $< \Psi_1/d$—a condition required to ensure that the breakdown occurs in the uniform portion of the field. (These are known as Rogowski electrodes.)

4.5 METHOD OF IMAGES

We have briefly introduced this method earlier in Sections 2.8.4 and 2.8.5 while discussing the field of an infinitely long line charge lying parallel to an infinite conducting plane. The method of images involves the conversion of an electric field to another equivalent field that is simpler to calculate. It is particularly useful for evaluating the fields of the point or the line charges near the boundaries which can be either conducting surfaces or dielectric surfaces. Thus this is a method of solving Laplacian and Poissonian field problems.

The above method was discovered by Lord Kelvin in 1848 for the purpose of solving problems in electrostatics in which the bounding surfaces between the media were to be accounted for. By using the principle of duality (i.e. the conjugate property of two fields), the method has been extended to the problems in magnetostatics and electromagnetics. However, at present we shall restrict our discussion to the electrostatic problems only. Lord Kelvin, in his analysis, showed that:

the field of an electric charge in front of a conducting sheet ≡ the field of the charge
+ the field of its mirror image in the conducting sheet

He used the term 'image' based on the similarity of this concept with the 'optical image'. But while the optical images are of two kinds, i.e. (a) real and (b) virtual, the electrostatic images are all virtual.

Thus the method of images implies, in electrostatics, determining an imaginary distribution of charges inside a conducting body (or even a dielectric body, as the case might be) in an electrostatic system, which would produce outside the body exactly the same field as that produced by the actual free charges over its surface; i.e. *the basic requirement of the method of images is that the effects of the boundaries shall be correctly represented by the images*. So now we consider a mathematical proof of this statement.

Proof: The electrostatic field in a region v can be represented by a scalar potential ϕ. Consider the Gauss' theorem,

$$\oiint_S \mathbf{A} \cdot \mathbf{n}\, dS = \iiint_v \nabla \cdot \mathbf{A}\, dv$$

where

A is an arbitrary vector in the region v

n is a unit vector drawn from within v, normal to S

S is the mathematical surface completely enclosing the volume v.

Now, let $\mathbf{A} = \psi \nabla \phi$, where ψ and ϕ are scalar functions in the region, then from the Gauss' theorem, we get

$$\oiint_S \psi \left(\frac{\partial \phi}{\partial n} \right) dS = \iiint_v \left\{ \psi \nabla^2 \phi + (\nabla \psi \cdot \nabla \phi) \right\} dv$$

This is the well known Green's first identity.

If $\psi = \phi$, then the above equation becomes

$$\oiint_S \phi \left(\frac{\partial \phi}{\partial n} \right) dS = \iiint_v \left\{ \phi \nabla^2 \phi + (\nabla \phi)^2 \right\} dv \tag{4.171}$$

This equation gives a relationship between the surface field and the volume field due to ϕ. We have to show that this relation is unique, if the image is to represent the boundary condition correctly.

So let us say that there are two solutions of ϕ, i.e. ϕ_1 and ϕ_2 both satisfying Eq. (4.171). Then both these potentials must satisfy the operating equation in v, i.e.

$$\nabla^2\phi_1 = -\frac{\rho}{\varepsilon_0} \qquad \text{and} \qquad \nabla^2\phi_2 = -\frac{\rho}{\varepsilon_0}$$

where ρ_C is the charge density within v.

$\therefore$ If $\chi = \phi_1 - \phi_2$, then $\nabla^2\chi = 0$.

Hence applying Eq. (4.171) to χ, we get

$$\oiint_S \chi\left(\frac{\partial\chi}{\partial n}\right) dS = \iiint_v \{(\nabla\chi)^2\}\, dv \tag{4.172}$$

If now, the left-hand side of this equation is zero, then the integrand of the right-hand side must also be zero. In such a case, $\nabla\chi = 0$, and both χ and $\phi_1 - \phi_2$ are constants. Thus the field derivable from ϕ_1 is identical with the field derived from ϕ_2. Therefore, the field is unique.

Equation (4.172) implies that the field will be unique if $\partial\chi/\partial n$ is zero everywhere on S. The condition for χ is that it should be constant, and not necessarily zero on S, because

$$\oiint_S \chi\left(\frac{\partial\chi}{\partial n}\right) dS = \chi \oiint_S \left(\frac{\partial\chi}{\partial n}\right) dS = \chi \iiint_v (\nabla^2\chi)\, dv \tag{4.173}$$

Thus the criterion for uniqueness is that either χ is constant or $\partial\chi/\partial n$ is zero everywhere on S. In physical terms this means that either the tangential or the normal components of the electric field be specified everywhere on S. A particularly important case which has attracted a lot of attention is when S is an equipotential surface. In such a case, the field is unique and in electrostatics the range of problems deal with the charge distributions near the perfect conductors (i.e. the enclosing surface S). However it should be noted that the condition that the enclosing surface S shall be a conducting surface is sufficient to determine the uniqueness of the field within v, but this is not a necessary condition. All that is necessary is that the value of the tangential or the normal electric field shall be specified at every point on the surface. (So the image methods can be applied to the dielectric boundaries as well.)

Thus from Eq. (4.172), the image problem in electrostatics can be formulated as follows:

(a) Consider a surface S which completely encloses v.
(b) Determine the tangential **E** field (or the normal field) on S due to the induced charges on the boundaries.
(c) Find a distribution of charges outside S which will give on S the same tangential (or normal) field.

Then this charge distribution would give the same field inside v, which would be given by the induced charges on the boundaries.

Note: The image distribution must be placed outside S, i.e. a virtual image in terms of optics. The image charges must be outside v because what is required is the field of these charges and not the charges themselves. Finally, although the field of the image distribution is unique, no such

uniqueness is required of the image distribution itself. It is not the uniqueness but the convenience that is the criterion of the image distribution.

4.5.1 Line Charge Parallel to the Surface of a Semi-infinite Dielectric Block

An infinite line charge Q C/m is in the region 1 of permittivity ε_1, at a distance a from the face of a semi-infinite slab of dielectric of permittivity ε_2 (Figure 4.17).

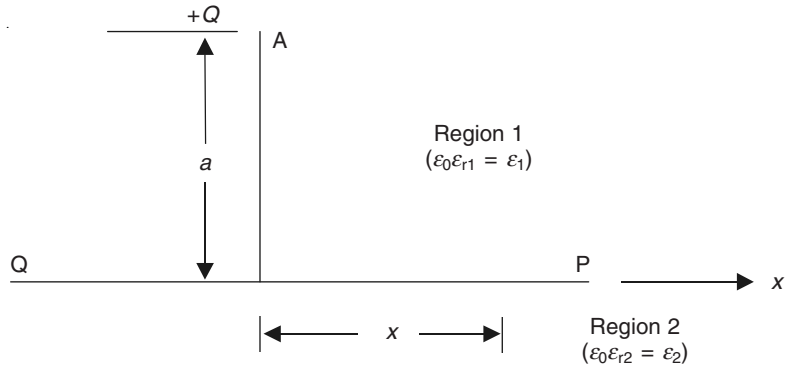


Figure 4.17 A line charge in front of a semi-infinite dielectric block.

Consider a point P on the interface as shown in the figure, and let the induced charge at P be σ. Therefore, the normal component of **E** (i.e. normal to the interface) at P due to the line charge Q is

$$E_{\mathrm{n}} = \frac{Qa}{2\pi\varepsilon_1(a^2 + x^2)} \tag{4.174}$$

This is the applied field.

To find the field due to the induced charge, we use the continuity condition for D_{n} on the interface, i.e.

$$\varepsilon_1\left(E_{\mathrm{n}} + E'_{\mathrm{n}}\right) = \varepsilon_2\left(E_{\mathrm{n}} - E'_{\mathrm{n}}\right)$$

where E'_{n} is the electric field due to the induced charges on the interface (Figure 4.18).

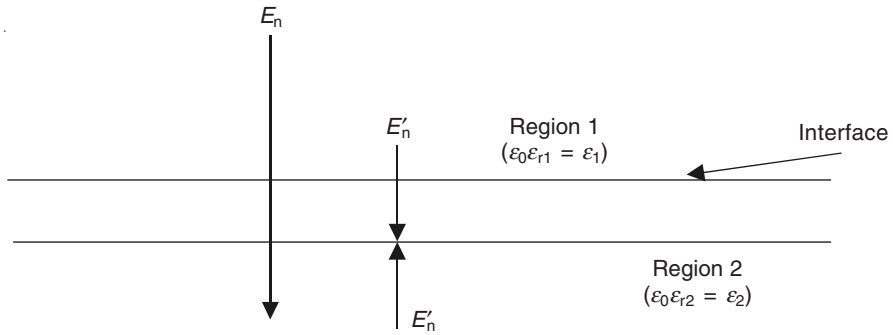


Figure 4.18 The induced field on the interface.

$$\therefore \qquad E'_{\mathrm{n}} = \left(\frac{\varepsilon_2 - \varepsilon_1}{\varepsilon_2 + \varepsilon_1} \right) E_{\mathrm{n}} \tag{4.175}$$

Thus the effect of the boundary (interface) on the field in the region 1 can be specified completely by the component E'_{n} along this interface and is given by the substitution

$$E'_{\mathrm{n}} = \left(\frac{Q}{2\pi\varepsilon_1} \right) \left(\frac{\varepsilon_2 - \varepsilon_1}{\varepsilon_2 + \varepsilon_1} \right) \left(\frac{a}{a^2 + x^2} \right) = \frac{\sigma}{2\varepsilon_1} \tag{4.176}$$

This is the same field distribution which would result from either (a) a charge $-Q\{(\varepsilon_2 - \varepsilon_1)/(\varepsilon_2 + \varepsilon_1)\}$ at a distance $y = a$ behind the interface line PQ in the region 2, or (b) equally from a charge $+Q\{(\varepsilon_2 - \varepsilon_1)/(\varepsilon_2 + \varepsilon_1)\}$ at a distance a from the interface in the region 1. Hence either of these charges can be used to account for the effect of the boundary. But only one can be used in the representation of the field in each of the two regions.

Thus considering first the field in the region 1, the choice of (b) would not satisfy the divergence of the flux condition (i.e. $\nabla \cdot \mathbf{D} = \rho_{\mathrm{C}}$) and hence cannot be used to represent the field in the region 1. But in the representation (a), the image charge lies outside the region 1 and so it does not affect the divergence of the flux there. This charge, called an image charge, gives, together with the source charge Q at A, the field solution for the region 1. (Note that this image charge does not apply to the field in the region 2.)

For the region 2, since the field here opposes the applied field, so the equivalent point charges giving the correct distribution of the field in this region will have opposite sign to the previous one and lie outside the region 2. So the required image is of the strength $-Q\{(\varepsilon_2 - \varepsilon_1)/(\varepsilon_2 + \varepsilon_1)\}$, lying at the point $y = +a$.

4.5.2 Point Charge Near an Infinite Grounded Conducting Plane

Consider a point charge Q at a distance D from an infinite conducting plane, which is earthed, as shown in Figure 4.19. The image charge of the charge Q in the conducting plane is a point charge $-Q$ placed symmetrically to Q with respect to the conducting plane. For, if the plane is removed

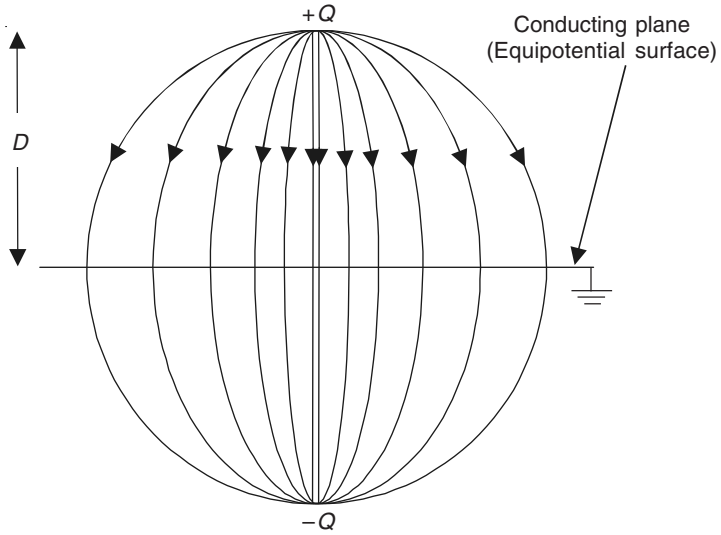


Figure 4.19 Image charge of the point charge Q near a plane conducting surface.

and $-Q$ is introduced instead, the field above the plane is the same as before (remembering that the field below the plane in reality does not exist). Hence the field of all the induced charges on the conducting plane can be reduced to the field of a single charge $-Q$ symmetrical to Q with respect to the conducting plane (Figure 4.20).

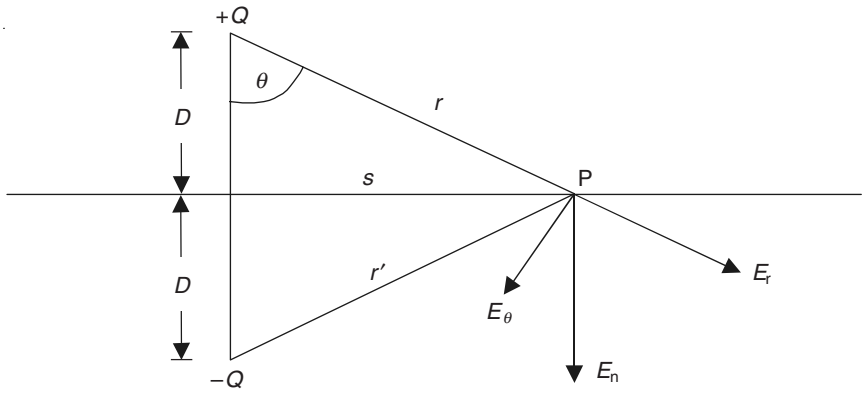


Figure 4.20 E_n on the conducting surface.

If the origin of the coordinate system is taken at Q, then the potential V at a point $P(r, \theta)$ will be

$$V = \left(\frac{1}{4\pi\varepsilon_0}\right)\left(\frac{Q}{r} - \frac{Q}{r'}\right) \tag{4.177}$$

where $r' = (r^2 + 4D^2 - 4rD\cos\theta)^{1/2}$.

The electric field intensity at P will be

$$E_r = -\left(\frac{\partial V}{\partial r}\right) = \left(\frac{1}{4\pi\varepsilon_0}\right)\left[\frac{Q}{r^2} - \frac{Q(r - 2D\cos\theta)}{r'^3}\right]$$

$$E_\theta = \left(-\frac{1}{r}\right)\left(\frac{\partial V}{\partial \theta}\right) = -\left(\frac{1}{4\pi\varepsilon_0}\right)\left(\frac{2QD\sin\theta}{r'^3}\right) \tag{4.178}$$

The lines of force are shown in Figure 4.19 and the equipotentials will be the curves orthogonal to the lines of force. The induced charge density σ' on the conducting surface is found from the normal component of the electric field intensity at the surface, i.e. $E_n = \sigma'/\varepsilon_0$. On the conducting surface, at all points $r = r'$, and

$$E_n = E_r\cos\theta - E_\theta\sin\theta = \frac{2QD}{4\pi\varepsilon_0 r^3} = -\frac{\sigma'}{\varepsilon_0}$$

$$\therefore \qquad \sigma' = \frac{QD}{2\pi r^3}$$

$\therefore$ The induced charge on the conductor surface is

$$= \int_0^\infty \sigma' 2\pi s\,ds = -\left(\frac{QD}{2}\right)\int_0^\infty \left\{\frac{2s\,ds}{(s^2 + D^2)^{3/2}}\right\} = -Q$$

as expected.

Note again that the image charge is always located outside the region where the field is being calculated. Also, it will be found that the force between the source charge and the conductor surface can be correctly found by calculating the Coulomb force between the source charge and its image. The force on Q in this case is $Q^2/\{(4\pi\varepsilon_0)\,(2D)^2\}$.

So far we have considered a single point charge near a conducting surface. The same reasoning would apply to any finite distribution of charges near such a surface. An important practical application of such a configuration, in its simplest form, is that of a two-wire transmission line running parallel to the earth. The diameter of the wire (= d) is much smaller than the distance between the wires (= D) and also their distance from the earth's surface (= h) (Figure 4.21).

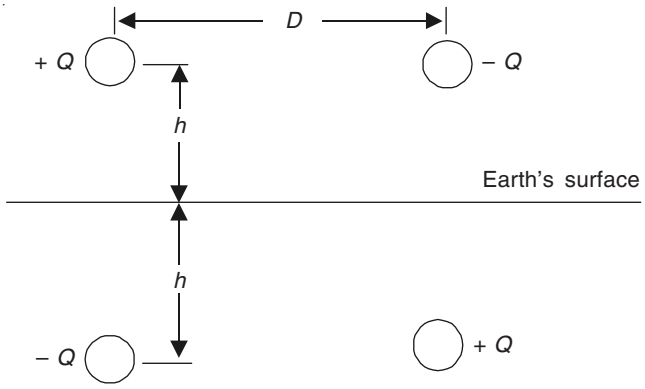


Figure 4.21 Two-wire transmission line parallel to the earth's surface and its image.

The problem posed is usually to find the capacitance per unit length of the system in presence of the earth's surface. The problem can then be approximated by considering the two infinite line charges $+Q$ and $-Q$ parallel to a conducting surface. These charges would induce charges on the surface the effects of which can be represented by taking account of the images $-Q$ and $+Q$ at a distance h from the surface in a direction opposite to the plane of the source lines. The problem can now be solved directly as it is an extension of the problem solved in Section 2.8.4. It can be easily checked that the capacitance per unit length comes out to be

$$\frac{\pi\varepsilon_0}{\ln\left[D/\left\{R\sqrt{1+(D/2h)^2}\right\}\right]} \qquad \text{where } 2R = d$$

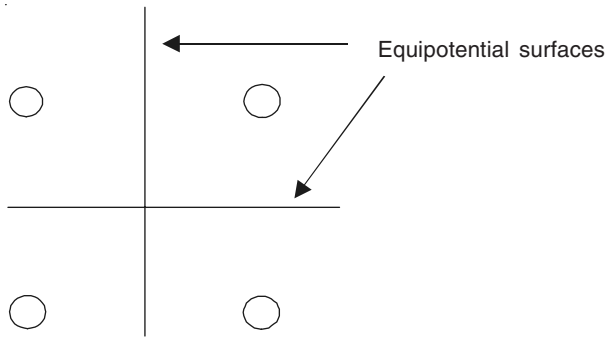


Figure 4.21(a) Image charges of a point charge in a 90° conducting corner.

Another problem of interest is that of a point charge placed near two conducting half-planes perpendicular to one another [Figure 4.21(a)]. The effects of the surfaces can be simulated by the three image charges as shown in Figure 4.21(a).

4.5.3 Line Charge Near a Circular Boundary

We consider a line charge Q per unit length, in a medium of permittivity $\varepsilon_1 (= \varepsilon_0 \varepsilon_{r1})$ running parallel to a circular cylinder of permittivity $\varepsilon_2 (= \varepsilon_0 \varepsilon_{r2})$ and radius R (Figure 4.22). Due to the

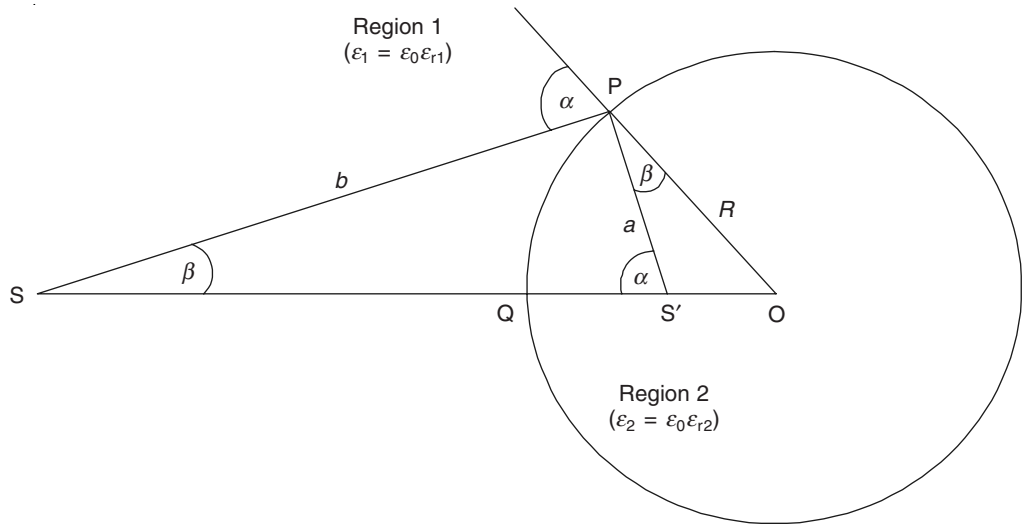


Figure 4.22 Line charge Q running parallel to the circular cylinder of radius R.

source charge Q at S, the normal component of the applied field strength E_n at a point P, as shown, on the circular boundary is

$$E_n = \frac{Q \cos \alpha}{2\pi \varepsilon_1 b} \tag{4.179}$$

where the angle α and the length b are as shown in Figure 4.22. Hence the normal component of the field due to the equivalent induced surface charge distribution is

$$E_n' = \left(\frac{\varepsilon_2 - \varepsilon_1}{\varepsilon_2 + \varepsilon_1} \right) \left(\frac{Q \cos \alpha}{2\pi \varepsilon_1 b} \right) \tag{4.180}$$

from Eq. (4.175).

From geometrical considerations, since S′ is the inverse point of S with respect to this circle, hence $OS \cdot OS' = OP^2 = OQ^2 = R^2$, and the angular relations are as shown in the figure. Considering the triangles PSS′ and PS′O, it can be shown that

$$a \cos \alpha + b \cos \beta = \frac{b \sin (\alpha + \beta)}{\sin \alpha}$$

and

$$a \sin \alpha = R \sin (\alpha - \beta)$$

On combining these two relationships, we get

$$\frac{\cos\alpha}{b} = \frac{\cos\beta}{a} - \frac{1}{R} \tag{4.181}$$

Hence the induced normal field E'_n can be rewritten as

$$E'_n = \left(\frac{\varepsilon_2 - \varepsilon_1}{\varepsilon_2 + \varepsilon_1}\right)\left(\frac{Q}{2\pi\varepsilon_1}\right)\left(\frac{\cos\beta}{a}\right) - \left(\frac{\varepsilon_2 - \varepsilon_1}{\varepsilon_2 + \varepsilon_1}\right)\left(\frac{Q}{2\pi\varepsilon_1}\right)\left(\frac{1}{R}\right) \tag{4.182}$$

This is the normal component of the field strength along the perimeter of the circular boundary due to the induced distributed charge. But this is also the normal component of the field due to a charge $-Q\{(\varepsilon_2 - \varepsilon_1)/(\varepsilon_2 + \varepsilon_1)\}$ at the inverse point S′ together with a charge $Q\{(\varepsilon_2 - \varepsilon_1)/(\varepsilon_2 + \varepsilon_1)\}$ at the centre of the circle.

For the region 2 inside the circle, the images producing E'_n must be outside the boundary and must produce a normal component of the field given by

$$E'_n = \left(\frac{Q}{2\pi\varepsilon_1}\right)\left(\frac{\cos\alpha}{b}\right)\left(\frac{\varepsilon_2 - \varepsilon_1}{\varepsilon_2 + \varepsilon_1}\right) \tag{4.183}$$

This requires an image charge of magnitude $-Q\{(\varepsilon_2 - \varepsilon_1)/(\varepsilon_2 + \varepsilon_1)\}$ and it would be located at the point S where the source charge was located.

Similarly it is possible to find the magnitudes and the positions of the images of a source line charge when it is located inside the circular boundary (i.e. at the point S′). The field inside the circle is given by an image of magnitude $Q\{(\varepsilon_2 - \varepsilon_1)/(\varepsilon_2 + \varepsilon_1)\}$ at the external inverse point S; and the field outside the circular boundary is due to an image of magnitude $-Q\{(\varepsilon_2 - \varepsilon_1)/(\varepsilon_2 + \varepsilon_1)\}$ at the same point as the source charge Q(i.e. the point S′) together with another image of magnitude $Q\{(\varepsilon_2 - \varepsilon_1)/(\varepsilon_2 + \varepsilon_1)\}$ at the centre O. This solution has been used to evaluate the fields inside electronic valves. This general solution can still be used when the cylindrical boundary is changed to an isolated conducting cylinder. In this case, in the above equations, we substitute $\varepsilon_2 \to \infty$. Then the images are of magnitudes $-Q$ at the inverse point and $+Q$ at the centre. The conducting cylinder, since it is isolated, acquires a potential, but has no charge on it because the flux from the charge from S passes to ∞, though some of it through the cylinder.

If, however, the conducting cylinder is earthed, its potential becomes zero. This causes a change in the image charge at the centre. Now all the flux from the source charge at S passes through the cylinder and it acquires a charge $-Q$. So the image charge at O must become zero and the image charge at S′ becomes $-Q$, and so the potential of the cylinder changes to zero.

4.5.4 Point Charge Near a Conducting Sphere

We consider first a grounded conducting sphere in which case it will be at zero potential. The source charge Q is positioned at the external point S whose distance from the centre of the sphere is a, the radius of the sphere being R (Figure 4.23). Since the sphere is grounded, it would be at zero potential, and so we have to find an image system, i.e. the magnitude and the position of the image charges, which together with the source charge would make the sphere to be at zero potential.

From the symmetry of the problem and our analysis in Section 4.5.3, it is obvious that it would be at the inverse point of S (say S′) on the line OS, satisfying the condition $OS \cdot OS' = R^2$. This can also be derived in a slightly different manner. Consider the diametral points P_1 and P_2 on

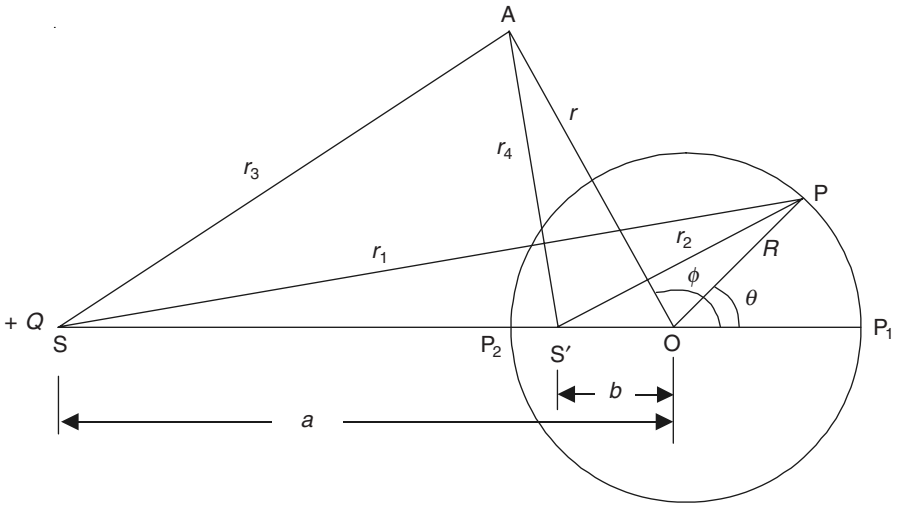


Figure 4.23 Point charge and the conducting sphere.

the sphere along the line OS, and assume that there is a charge Q' at S′ also on the line OS which together with Q at S make the circular boundary at zero potential. Then,
for P_1,

$$\frac{Q}{a+R}+\frac{Q'}{b+R}=0$$

and for P_2,

$$\frac{Q}{a-R}+\frac{Q'}{R-b}=0 \tag{4.184}$$

Solving these two equations, we have

$$Q'=-\left(\frac{R}{a}\right)Q \quad \text{and} \quad b=\frac{R^2}{a}, \text{ i.e. the inverse points} \tag{4.185}$$

We still have to find out if this charge arrangement would make the whole spherical surface to be at zero potential. So we take a general point P on the sphere, at which

$$4\pi\varepsilon_0 V=\frac{Q}{r_1}+\frac{Q'}{r_2} \tag{4.186}$$

where

$$r_1=\sqrt{a^2+R^2+2aR\cos\theta}$$

and

$$r_2=\sqrt{b^2+R^2+2bR\cos\theta} \tag{4.187}$$

Substituting for Q' and b from Eq. (4.185), we get $V = 0$, at P. Substituting for Q' and b, we get the potential V as

$$V=\left(\frac{1}{4\pi\varepsilon_0}\right)\left(\frac{Q}{r_1}-\frac{Q(R/a)}{r_2}\right)=0$$

where

$$r_1 = \sqrt{a^2 + R^2 + 2aR\cos\theta}$$

and

$$r_2 = \sqrt{\left(R^2/a\right)^2 + R^2 + 2\left(R^3/a\right)\cos\theta} \tag{4.188}$$

If now we consider an arbitrary point A, anywhere in the plane as shown in the figure, then the potential at A(r, ϕ) is

$$V = \left(\frac{1}{4\pi\varepsilon_0}\right)\left(\frac{Q}{r_3} - \frac{Q(R/a)}{r_4}\right)$$

where

$$r_3 = \sqrt{a^2 + r^2 + 2ar\cos\phi}$$

and

$$r_4 = \sqrt{\left(R^2/a\right)^2 + r^2 + 2\left(R^2/a\right)r\cos\phi} \tag{4.189}$$

To find the electric field intensity,

$$\mathbf{E} = -\nabla V = -\left(\frac{\partial V}{\partial r}\right)\mathbf{i}_r, \text{ from the symmetry of the problem.}$$

$$\therefore \qquad E_r = \left(\frac{1}{4\pi\varepsilon_0}\right)\left[\frac{Q(r + a\cos\phi)}{r^2} - \frac{RQ\left\{r + (R^2/a)\cos\phi\right\}}{r_4^3 a}\right] \tag{4.190}$$

and at $r = a$

$$E_r = \left(-\frac{1}{4\pi\varepsilon_0}\right)\left[\frac{Q(a^2 - R^2)}{R(a^2 + r^2 + 2ar\cos\phi)^{3/2}}\right]$$

$$= \frac{\sigma'}{\varepsilon_0} \tag{4.191}$$

where σ' is the induced charge density on the spherical surface.

$\therefore$ The total induced charge on the sphere is

$$Q' = \int_0^{\pi} \sigma' 2\pi a^2 \sin\theta\, d\theta = -\frac{QR}{a} \tag{4.192}$$

which is the same as the image charge. This is true because of the Gauss' theorem. [This is an extension of the "Weiss' Sphere Theorem" of images in Fluid Dynamics (refer *Textbook of Fluid Dynamics* by F. Chorlton).] Next we consider a point charge near a charged conducting sphere. When the potential of the sphere is different from zero, it is still possible to use the method of images to solve the problem. If Q is the point charge at the source point, first we replace the sphere by an image charge $Q' = (QR/a)$ at the inverse point of the source point, i.e. at the point (R^2/a) away from the centre of the sphere on the line joining the centre of the sphere to the source point, and to that we add another image charge $(Q_S - QR/a)$ located at the centre of the sphere, where Q_S is the charge on the sphere. The surface charge density on the spherical boundary is then $\sigma' + \sigma''$, where σ' is the charge distribution calculated from Q and Q', and σ'' is the charge density due to $(Q_S - Q')$.

Some general comments. So far we have considered only some typical electrostatic field problems solved by the method of images. We shall see further applications of this method when we study the problems in magnetostatics as well as electromagnetism. However it would be obvious from what we have studied so far that the method of images is an alternative way of finding the Green's function for the boundary surfaces in Laplacian and Poissonian field problems.

PROBLEMS

4.1 The electric potential distribution in a metal strip of uniform thickness and constant width a, and extending to infinity from $y = 0$ to $y \to \infty$ is obtained as

$$V = V_0 \exp\left(-\frac{\pi y}{a}\right) \sin\left(\frac{\pi x}{a}\right)$$

Show the coordinate system (with reference to the plate) used and find the boundary conditions used for the above potential distribution.

4.2 A high voltage coaxial cable consists of a single conductor of radius R_i, and a cylindrical metal sheath of radius R_o ($R_o > R_i$) and a homogeneous insulating material between the two. Since the cable is very long compared with it diameter, the end effects can be neglected, and hence the potential distribution in the dielectric can be considered to be independent of the position along the cable. Write down the Laplace's equation in the circular–cylinder coordinates, and state the boundary conditions for the problem. Solving the Laplace's equation, show that the potential distribution in the dielectric is

$$V = \frac{V_s \ln (R_o/r)}{\ln (R_o / R_i)}$$

and hence find the capacitance of the cable per unit length. Note that V_s is the applied (supply) voltage on the inner conductor of radius R_i.

4.3 A polystyreme cylinder of circular cross-section has a radius R and axial length $2d$. Both ends are maintained at zero potential. An electric potential V is applied on the cylindrical surface as

$$V = V_0 \cos (\pi z/2d)$$

Obtain an expression for the potential at any point in the material.

(*Hint:* Take the origin of the coordinate system at the mid-point between the ends.)

4.4 A line charge of Q_1 per unit length runs parallel to a grounded conducting corner (right-angled) and is equidistant (= a) from both the planes. Show that the resultant force on the line charge is

$$\frac{-Q_1^2}{\left(4\sqrt{2}\pi\varepsilon_0 a\right)} \text{ per unit length,}$$

and is directed along the shortest line joining the line charge with the corner.

4.5 Using Eqs. (4.182) and (4.190) discuss the salient points of difference, when a line charge (say Q per unit length) is placed at a distance b from the centre of an infinitely long conducting cylinder of radius R, and a point charge (say Q) placed at the same distance from the centre of a conducting sphere of radius R. (The medium being air, its permittivity is ε_0.)

5 Approximate Methods of Solving Electrostatic Field Problems

5.1 INTRODUCTION

In the preceding chapters, several methods of solving the electrostatic field problems analytically have been explained. The solutions of such problems have been either exact or almost exact. But in most of these problems, the geometry of the conducting or the dielectric bodies has been relatively simple; the boundaries were planes, spheres, cylinders, corners, etc.

However, in practice, a significant number of electrostatic systems have such complicated geometry that an analytical solution is not feasible. For a rough estimate, in some cases it is possible to approximate a real system by some geometrically simpler system which can then be solved by one of the analytical methods. But for a more precise analysis, it is preferable to use some approximate method for solving such problems. Now we shall discuss in some details some of such methods and also explain the logic of some in brief. First we shall discuss the graphical method and later the semi-graphical method, both of which are mostly applicable to two-dimensional problems.

5.2 GRAPHICAL METHOD OF SOLVING ELECTROSTATIC PROBLEMS

Electrostatic fields and many other physical problems (such as magnetic fields, temperature, and velocity) have streamline properties in the space between the surfaces between which the potential differences exist. In such fields, lines can be drawn to indicate by their direction and spacing, the direction and the intensity of the flow in the region. Other lines, orthogonal to these flow-lines, connect the points in the field having equal potentials. There is no component of flow along an equipotential (Figure 5.1). Such streamline flow satisfies the 'continuity equation'. The 'equation of continuity' is a mathematical formulation of the principle that the rate of accumulation of a quantity in a region is equal to the net rate at which the material crosses the walls of the region. Thus for an incompressible fluid in a region with no sources and sinks, the total flux through a closed surface must vanish.

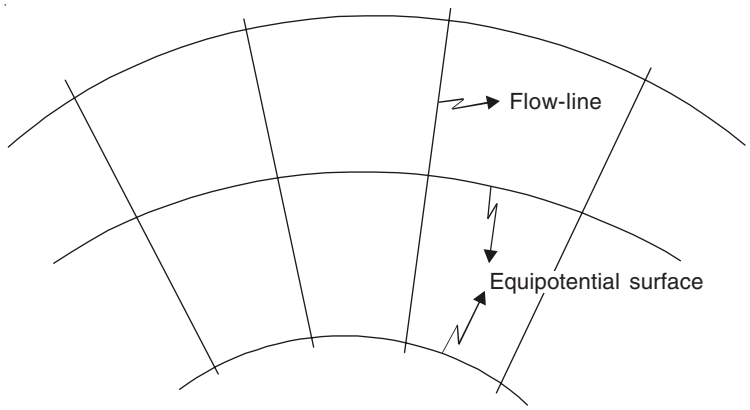


Figure 5.1 Flow-lines and equipotentials.

The properties of a field can be defined completely if the equipotential surfaces and the lines of flux or flow can be mapped. When these two sets of lines, i.e. the equipotentials and the flux-lines, are orthogonal, then the field is Laplacian. This statement can be proved mathematically and since a very large number of electrostatic fields are Laplacian in nature, we shall give the proof of this statement before we proceed with further details of the method under discussion. However, there is a further point which needs to be explained before we consider this proof. When these sets of orthogonal lines intersect, the geometrical figures obtained are curvilinear rectangles, or in special cases curvilinear squares.

5.2.1 A Note on Curvilinear Squares

A curvilinear square is indicated in Figure 5.2. It may be defined as an area bounded by two pairs

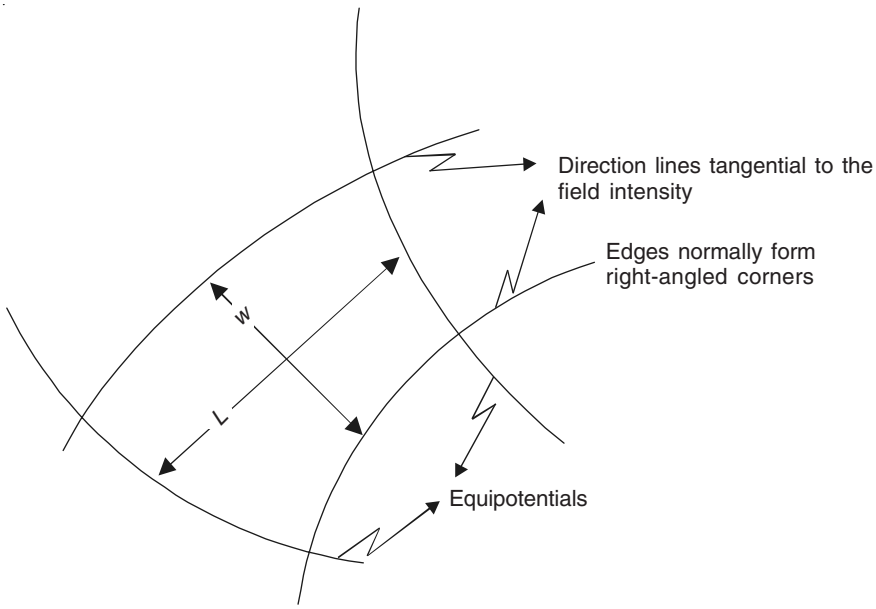


Figure 5.2 A curvilinear square.

of curves, belonging to two mutually orthogonal families, which on further sub-division become rectilinear squares in the limit. In the present case, one set of curves correspond to the direction of the field intensity in a given situation, and the other orthogonal set of curves to equipotential surfaces. As the equipotential surfaces are perpendicular to the direction of the field intensity lines in the region, the corners of the curvilinear squares are normally at right angles. A curvilinear square is always drawn so that its average length L equals the average width w. (More generally in a curvilinear rectangle, the ratio L/w will be different from unity.) The precise length and the width of the square are difficult to define but the experience in drawing curvilinear squares develops a feel for the appearance of a properly drawn square. A test which may be applied to a figure to determine if it is a curvilinear square is to sub-divide the figure into smaller curvilinear squares. In the limit as the curvilinear squares become smaller, each such square should approach a perfect square. If they do not, then the original figure is not a curvilinear square. This implies that the process of drawing and checking can be quite long, but with practice it is possible to draw fairly good graphs of fields by this method in a reasonably short time. For approximate answers, the graph need not be perfect. For calculation purposes, N_f is taken as the number of curvilinear squares along each flux line (from the +ve electrode to the –ve electrode) and N_e is the number of squares along a 'complete' equipotential (and not one which goes to infinity).

5.2.2 A Proof that the Potential Associated with a Field-plot Consisting of Curvilinear Rectangles Satisfies the Laplace's Equation

Consider two scalar functions Φ and Θ of two variables x and y such that the curves of constant Φ and constant Θ form the curvilinear rectangles as shown in Figure 5.3.

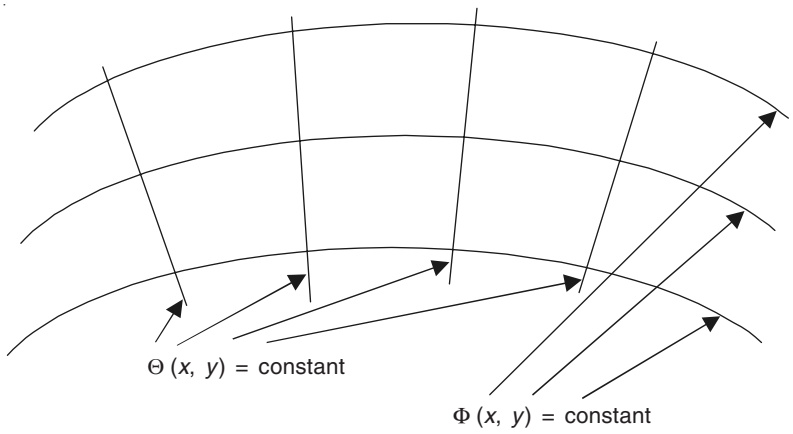


Figure 5.3 Field of curvilinear rectangles.

The statement that the curves of constant Φ and constant Θ form the curvilinear rectangles has the following two implications:

(a) Since the rectangles are formed by the constant curves, it means that Φ and Θ are orthogonal everywhere, i.e. the vector normal to the curves of constant Φ is normal to the vector normal to the curves of constant Θ. Or (grad Φ) is perpendicular to (grad Θ) at every point.

$$\therefore \quad (\nabla\Phi)\cdot(\nabla\Theta) = 0 \tag{5.1}$$

In the two-dimensional rectangular Cartesian coordinates, this relation becomes

$$\left(\mathbf{i}_x \frac{\partial\Phi}{\partial x} + \mathbf{i}_y \frac{\partial\Phi}{\partial y}\right)\cdot\left(\mathbf{i}_x \frac{\partial\Theta}{\partial x} + \mathbf{i}_y \frac{\partial\Theta}{\partial y}\right) = 0$$

or

$$\frac{\partial\Phi}{\partial x}\frac{\partial\Theta}{\partial x} + \frac{\partial\Phi}{\partial y}\frac{\partial\Theta}{\partial y} = 0 \tag{5.2}$$

(b) The length to width ratio of the rectangles is constant throughout the x–y plane. This means that the changes in Φ and Θ between the curves are related by a constant ratio. In terms of the gradient operator, this relation is

$$(\nabla\Phi)\cdot(\nabla\Phi) = K\{(\nabla\Theta)\cdot(\nabla\Theta)\} \tag{5.3}$$

where K is a +ve scale operator.

$\therefore$ In the rectangular Cartesian system in two-dimensions, Eq. (5.3) becomes

$$\left(\frac{\partial\Phi}{\partial x}\right)^2 + \left(\frac{\partial\Phi}{\partial y}\right)^2 = K\left\{\left(\frac{\partial\Theta}{\partial x}\right)^2 + \left(\frac{\partial\Theta}{\partial y}\right)^2\right\} \tag{5.4}$$

From Eq. (5.2),

$$\frac{\partial\Phi}{\partial y} = -\frac{\dfrac{\partial\Phi}{\partial x}\dfrac{\partial\Theta}{\partial x}}{\dfrac{\partial\Theta}{\partial y}} \tag{5.5}$$

Substituting in Eq. (5.4) and rearranging

$$\left(\frac{\partial\Phi}{\partial x}\right)^2 = \left\{1 + \frac{\left(\dfrac{\partial\Theta}{\partial x}\right)^2}{\left(\dfrac{\partial\Theta}{\partial y}\right)^2}\right\} = K\left\{\left(\frac{\partial\Theta}{\partial x}\right)^2 + \left(\frac{\partial\Theta}{\partial y}\right)^2\right\}$$

$$\therefore \quad \left(\frac{\partial\Phi}{\partial x}\right)^2 = K\left(\frac{\partial\Theta}{\partial y}\right)^2 \tag{5.6}$$

Again, rewriting Eq. (5.2) as

$$\frac{\partial\Phi}{\partial x} = -\frac{\dfrac{\partial\Phi}{\partial y}\dfrac{\partial\Theta}{\partial y}}{\dfrac{\partial\Theta}{\partial x}} \tag{5.7}$$

and then substituting in Eq. (5.4), we derive that

$$\left(\frac{\partial\Phi}{\partial y}\right)^2 = K\left(\frac{\partial\Theta}{\partial x}\right)^2 \tag{5.8}$$

Differentiating Eqs. (5.6) and (5.8) with respect to x and y, respectively, we get

$$2\frac{\partial\Phi}{\partial x}\frac{\partial^2\Phi}{\partial x^2} = 2K\frac{\partial\Theta}{\partial y}\frac{\partial^2\Theta}{\partial y\,\partial x}$$

and

$$2\frac{\partial\Phi}{\partial y}\frac{\partial^2\Phi}{\partial y^2} = 2K\frac{\partial\Theta}{\partial x}\frac{\partial^2\Theta}{\partial x\,\partial y}$$

Since

$$\frac{\partial^2\Theta}{\partial y\,\partial x} = \frac{\partial^2\Theta}{\partial x\,\partial y}, \text{ from the above two equations,}$$

$$\frac{\dfrac{\partial\Phi}{\partial x}\dfrac{\partial^2\Phi}{\partial x^2}}{\dfrac{\partial\Theta}{\partial y}} = \frac{\dfrac{\partial\Phi}{\partial y}\dfrac{\partial^2\Phi}{\partial y^2}}{\dfrac{\partial\Theta}{\partial x}}$$

or

$$\frac{\partial^2\Phi}{\partial x^2} - \frac{\dfrac{\partial\Phi}{\partial y}\dfrac{\partial\Theta}{\partial y}}{\dfrac{\partial\Phi}{\partial x}\dfrac{\partial\Theta}{\partial x}}\frac{\partial^2\Phi}{\partial y^2} = 0 \tag{5.9}$$

But from Eq. (5.2)

$$\frac{\dfrac{\partial\Phi}{\partial y}\dfrac{\partial\Theta}{\partial y}}{\dfrac{\partial\Phi}{\partial x}\dfrac{\partial\Theta}{\partial x}} = -1 \tag{5.10}$$

Hence substituting Eq. (5.10) in Eq. (5.9), we get

$$\frac{\partial^2\Phi}{\partial x^2} + \frac{\partial^2\Phi}{\partial y^2} = 0 \tag{5.11}$$

which is the Laplace's equation in two-dimensions.

Similarly it can also be shown that Θ also satisfies the Laplace's equation.

5.2.3 Plotting Technique

As stated earlier, for a complete definition of a field we need to have the equipotentials and the lines of flux (or flow). The flux function is vectorial having a direction of flow and the strength in flow units per area (flux density, say). The potential is a scalar function having only magnitude at any point. These names '*flux lines*' and 'equipotentials' have been given to two sets of orthogonally intersecting lines on physical grounds and every field map can be used to solve two problems—the given problem and its dual—by simply interchanging the names of the two sets of lines; that is, the equipotentials of one problem become the lines of force in the other, and conversely. This conjugate property of the two fields is an example of the 'Principle of Duality'. (A detailed discussion of this principle is given in Appendix 1). Based on the mathematical theory discussed so far, the complete solution of the field would be obtained if a set of equipotential lines and flux lines are drawn to form a continuous system of curvilinear rectangles

between the boundary electrodes. However it is more convenient in sketching to make the curvilinear rectangles into the curvilinear squares.

The procedure is to draw one or more equipotential lines to divide the area suitably between the boundaries, and then draw in the flux lines to cross the equipotential lines at right angles, and form curvilinear squares as far as possible. The first trial usually does not give curvilinear squares throughout the region, and the lines are corrected by trial and error till such squares are obtained over the whole area. Figure 5.4 shows such a trial solution. The lines are usually started between

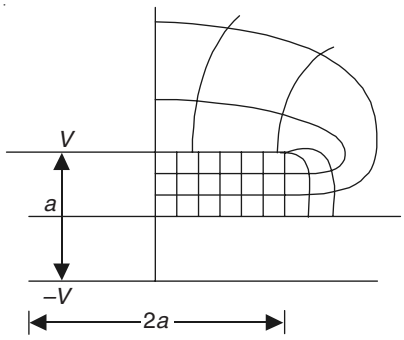


Figure 5.4 Curvilinear squares for an electrode system.

the parallel portions of the boundaries where the field would be expected to be uniform. For the space between the electrode system in which the curvilinear squares are drawn, they satisfy the following conditions:

(a) All the equipotentials and the flux lines intersect at right angles.
(b) For any curvilinear square, the mean width w between the flux lines is equal to the mean length L between the equipotentials.
(c) All the flux lines meet the boundary electrode normally.
(d) The large curvilinear squares can be sub-divided to produce smaller curvilinear squares.

The drawing is thus a complete solution of the field and the curvilinear squares have the following quantitative properties, assuming a unit depth of the field perpendicular to the plane of the paper:

(a) The potential difference between the equipotentials across all the curvilinear squares is the same.
(b) Equal quantities of flux pass through all the volumes formed by the curvilinear squares.
(c) The capacitance, the permeance, the resistance, etc. of each curvilinear square is the same.
(d) The energy of each curvilinear square volume is the same.
(e) The flux density and the electric field strength vary throughout the field inversely as the size of the squares.

From the plot, we can calculate the capacitance of an electrode system, i.e. if V is the potential difference between the electrodes and ΔV is the potential difference across the curvilinear square, then $V = N_f \Delta V$, where N_f is the number of squares along a flux line, from the +ve electrode to the –ve electrode.

$\therefore$ The electric field intensity **E** is given by

$$E \simeq \frac{\Delta V}{L}$$

and hence the magnitude of the electric displacement vector **D** is

$$D = \varepsilon E \simeq \frac{\varepsilon \Delta V}{L}$$

The flux Ψ_D of the displacement vector **D** (or the flux density vector) represented by the successive flow lines in the figure, is

$$\Delta \Psi_D \simeq Dw \simeq \varepsilon \Delta V \frac{w}{L}$$

Hence, from the generalized form of the Gauss' theorem, the charge Q_l on the conductor per unit length is

$$Q_l = N_e \, \Delta \Psi_D = N_e \, \varepsilon \Delta \mathrm{V} \frac{w}{L} = \varepsilon V \left(\frac{N_e}{N_f} \right) \left(\frac{w}{L} \right)$$

where N_e is the number of curvilinear squares along a complete equipotential, and hence is the number of lines of the displacement vector emerging from the contour of the conductor. Hence the capacitance, per unit length, between the parallel plates of the capacitor of Figure 5.4 is given by

$$C_l = \frac{Q_l}{V} = \varepsilon \left(\frac{N_e}{N_f} \right) \left(\frac{w}{L} \right) = \varepsilon \left(\frac{N_e}{N_f} \right), \text{ when } w = L \text{ (as for the curvilinear squares).}$$

For the capacitor shown in the figure, $N_e = 21$, $N_f = 6$, and the gap between the plates = a, the width of the plates = $2a$, and the dielectric is air, i.e. $\varepsilon = \varepsilon_0 = 8.854 \times 10^{-12}$.

$$\therefore \qquad C_l = 8.854 \times 10^{-12} \left(\frac{21}{6} \right) = 31 \text{ pF/m}$$

whereas using the formula which neglects the fringing effect,

$$C_{l\text{th}} = \varepsilon \frac{A}{d} = 8.854 \times 10^{-12} \left(\frac{2a}{a} \right) = 17.71 \text{ pF/m}$$

To find the degree of precision of this method, it should be appreciated that the error reduces for the smaller curvilinear squares. But even with moderate size squares, the error can be restricted to 10 to 15%. This can be checked for this problem by comparing the rigorous solution obtained by the method of conformal transformation.

5.2.4 Two-dimensional Multi-dielectric Fields

So far we have discussed those electrostatic fields which have either air or a single dielectric in the gap region. But when on an interface boundary, the permittivity changes from ε_1 to ε_2, there would be refraction of the flux lines as well as of the equipotentials, and the capacitance of the curvilinear squares in the changed dielectric would suddenly change in the proportion of the permittivities. Since we wish to keep the capacitance of each element to be the same, the length to width ratio must be changed in the changed dielectric and hence the curvilinear squares become curvilinear rectangles. For the unit depth of the field where L is the mean length between the equipotentials and w the mean width between the flux lines, the capacitance of any curvilinear rectangle is proportional to $\varepsilon w/L$. If in the dielectric of permittivity ε_1, the curvilinear squares have been drawn, i.e. $w/L = 1$, then in the dielectric of permittivity ε_2, the curvilinear

rectangles must be drawn with $w/L = \varepsilon_1/\varepsilon_2$. At the dielectric boundary, the tangential components of the field strength E_t and the normal components of the flux density D_n are continuous. But the normal components of the field strength E_n and the tangential components of the flux density D_t are discontinuous. Hence, we get (Figure 5.5)

$$D_{n1} = D_{n2}, \quad E_{t1} = E_{t2}, \quad D_{t1} = D_{t2}\left(\frac{\varepsilon_1}{\varepsilon_2}\right), \quad E_{n1} = E_{n2}\left(\frac{\varepsilon_1}{\varepsilon_2}\right)$$

$$\frac{\tan\alpha_1}{\tan\alpha_2} = \frac{D_{t1}/D_{n1}}{D_{t2}/D_{n2}} = \frac{D_{t1}}{D_{t2}} = \frac{\varepsilon_1 E_{t1}}{\varepsilon_2 E_{t2}} = \frac{\varepsilon_1}{\varepsilon_2} \tag{5.12}$$

For constructing a plot, it is best to obtain the refraction effect by setting off equal normal and tangential components proportional to ε, at the point of intersection of the flux density line with the dielectric interface as shown in Figure 5.5. The equipotential tangents are constructed at 90° to the flux line tangents.

While plotting the multi-dielectric fields the whole area should be divided into coarse

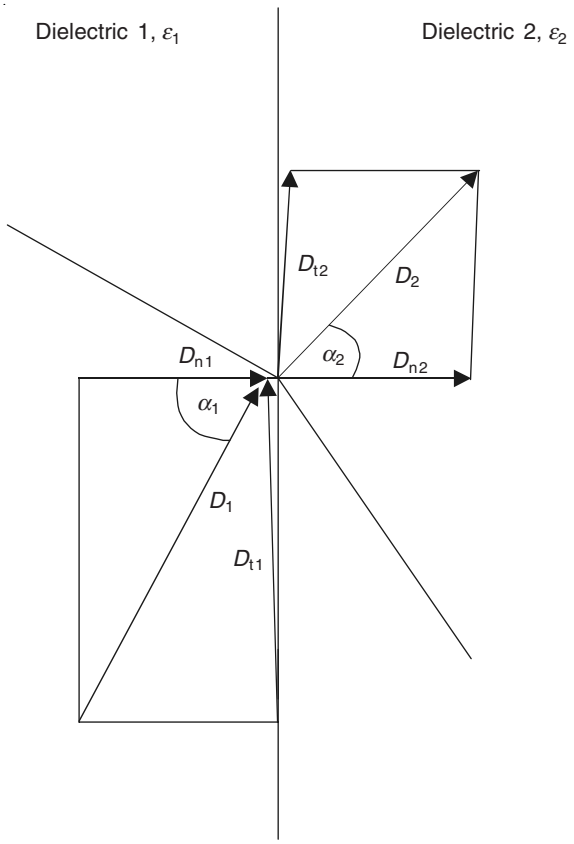


Figure 5.5 Discontinuity at the interface.

curvilinear rectangles, noting that the flux lines entering the region of higher permittivity always bend towards the interface surface whereas the equipotentials bend away from the interface.

General comments. Axi-symmetric fields can also be plotted, but the work involved is so much that it is rarely justified. The interested readers should refer to *Introductory Engineering Electromagnetics* by B.D. Popovic, and the paper "Graphical Field Plotting" by E.G. Wright, *Electrical Review*, March 1957. So far we have discussed only the Laplacian fields, but it is also possible to analyze the Poissonian fields containing the regions of source charges. We shall consider such problems by analogy when we take up the problems of magnetic fields in current-carrying regions.

5.3 EXPERIMENTAL METHODS

Since the analytical expressions for the field quantities are possible mostly for simple geometries, many experimental methods have been developed for problems with complex geometries. These experimental methods usually avoid the rigour of the advanced mathematical techniques, but have great value in helping the visualization of the field distributions and as checks on the analytical solutions. For quantitative mapping of the electrostatic fields, it is usually sufficient to obtain a plot of the equipotentials or of the flux lines from which the other set can be obtained graphically by the method of curvilinear squares. Again these methods are mostly used for the two-dimensional and the axi-symmetric problems, since one single section gives all the required information. For the general three-dimensional problems, more than one section has to be considered, and also further careful interpretation of the individual plots to produce the actual field picture.

As we proceed with our study of electromagnetism, we shall see that there are other field phenomena apart from the electrostatics (i.e. magnetostatics, electric current, etc.) which show the same basic relationships between the characteristic field vectors, and hence close analogies can be established between these different fields. Such analogies form the basis of these experimental methods because easier substitutions can be made where the original field is difficult to analyze. Electric current flow analogy is the most commonly used technique. The choice of the medium, in which the current flow is studied, decides the technique. Some media are more commonly used for studying the electrostatic problems while other media are preferred for the magnetic field problems. We shall discuss such techniques in the relevant chapters of our present study. However it should be borne in mind that even though a particular technique has been preferred for analyzing a particular type of field, it does not preclude its (the technique's) use for other types of field problems, so long as the field under consideration satisfies the same operating equation, i.e. say the Laplace's equation or the Poisson's equation. We shall now discuss the principles of some of these experimental methods, keeping in mind that those methods which are more commonly used for solving other types of fields—say, magnetic—will be described in the relevant sections.

5.3.1 Electrolytic Tank Method

A very general use of the analogy of the electric fields (or currents) in the conducting medium to the electrostatic fields or any potential fields, has been by means of the electrolytic current distributions (i.e. in a conducting liquid) either at direct voltages or at lower audio frequencies where the magnetic induction effects can be neglected. The arrangement consists of a large tank made of glass or impregnated wood lined with copper or lava slabs, filled with distilled water and a small amount of spring or well water to obtain a proper degree of conductivity. Even ordinary tap water can be used or even a very weak solution of copper sulphate. The electrodes, which are usually made of copper, are dipped in the electrolyte, and a probe, usually a short piece of nickel

or platinum wire of about 0.2 mm diameter, is used for indicating the local potential. The probe should be insulated over its entire length, except for a very small length (say, about 1 cm) at its extreme end. The probe can be sealed in glass and there should be a metal sheath on its outside, so that the sheath acts as a shield.

The electric circuit (see Figure 5.6) is essentially a Wheatstone bridge. Its two arms are

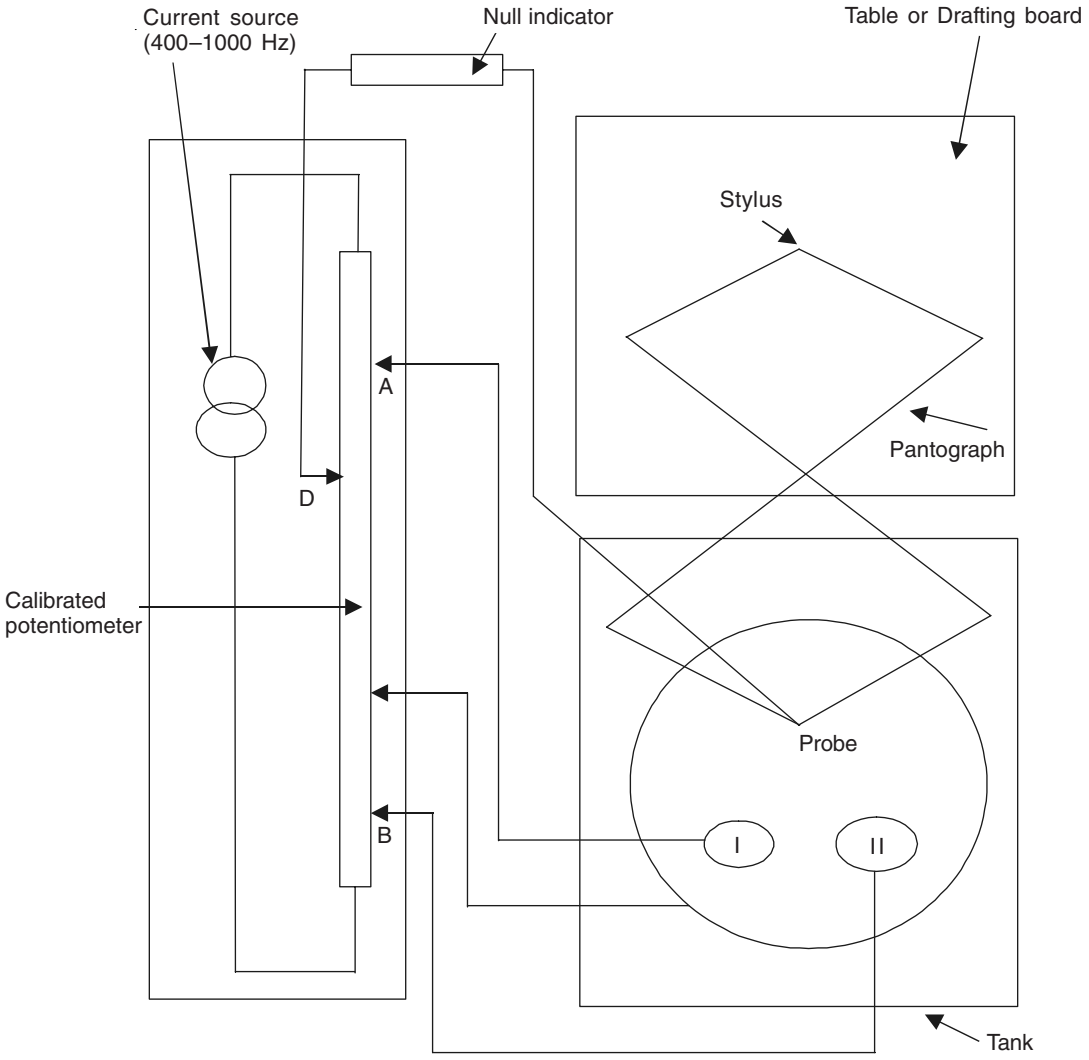


Figure 5.6 Electrolytic tank for the experimental solution of the Laplace's equation.

formed by the probe and the electrodes I and II, and the other two arms are AD and BD on the calibrated potentiometer. A selected value of the potential is set on the tap D of the potentiometer and the probe is moved till its potential equals to that of D, as would be indicated by the null indicator. The position of the probe is transmitted to a stylus resting on an adjoining table. The movement of the stylus is controlled either by a *x-y* motion carriage system fixed to the rim of the tank, or by a pantograph as shown in the figure. Usually a complete equipotential can be drawn for a fixed position of D. Automatic plotting of all required equipotentials can be achieved by

driving the probe at constant speed in one direction and adjusting its position in the orthogonal direction by a suitable servomechanism which would correct to the zero difference, the probe potential with respect to the potential line to be drawn. At the end of each travel, the servo-mechanism has to select the position of the probe for the next equipotential. Once all the equipotentials are obtained, the flow lines can be drawn by the method of curvilinear rectangles.

The electrolytic tank was first demonstrated by Fortescue in 1913, who used direct voltage. The disadvantage of polarization led to the modification of using alternating voltages from low power frequencies up to 500 Hz. Higher frequencies up to 1.5 kHz have been used, but at higher frequencies, increasing capacitive effects create difficulties in the zero balance. It is very important to keep the electrode surfaces very clean because even slight oxidation can increase the local surface resistance.

The advantage of the electrolytic tank method is that it is also capable of reproducing a wide range of three-dimensional field distributions in a uniform medium. Even the effects of two different dielectric media in a system has been studied and represented on the tank. For a detailed discussion of the method, the reader is referred to Weber, Moon and Spencer.

5.3.2 Conducting Paper Analogue

This method has been very widely used for the magnetic field problems, even though it is equally capable of solving the two-dimensional electrostatic field problems, as this method is also based on current-flow analogy and hence can solve all Laplacian field problems. We shall discuss this method in detail, with its mathematical basis, during our study of the solutions of the magnetic field problems.

5.3.3 Elastic Membrane Method (Rubber Sheet Analogy)

This is another method of solving the Laplacian field problems in two-dimensions by using a rubber sheet stretched over the supports which represent the boundary conditions. The rubber sheet then takes the form of a surface in three-dimensional space in which the height represents the potential ϕ as a function of x and y. The exact partial differential equation governing the vertical displacement ϕ of a horizontal sheet is

$$\left(\frac{\partial^2\phi}{\partial x^2}\right)\left[1+\left(\frac{\partial\phi}{\partial y}\right)^2\right]+\left(\frac{\partial^2\phi}{\partial y^2}\right)\left[1+\left(\frac{\partial\phi}{\partial x}\right)^2\right]-\left(\frac{\partial^2\phi}{\partial x\,\partial y}\right)\left(\frac{\partial\phi}{\partial x}\right)\left(\frac{\partial\phi}{\partial y}\right)=0 \tag{5.13}$$

If the displacements are small, then the slopes of the deformed membrane would be quite small, i.e. $(\partial\phi/\partial x) << 1$ and $(\partial\phi/\partial y) << 1$, and these can be neglected. Therefore, Eq. (5.13) reduces to

$$\frac{\partial^2\phi}{\partial x^2}+\frac{\partial^2\phi}{\partial y^2}=0$$

and so the displacement ϕ in such cases approximately satisfies the Laplace's equation. If the tangent planes at any point are kept within 15° of the horizontal plane, then the above conditions are generally satisfied. So the models have to be rather large, but of small height differences. Surgical rubber sheet of about 0.1 cm thickness can be conveniently used for such models. The sheet is spread over the electrode surfaces and pulled over a wooden frame and fastened to it. For further details, the reader is referred to Weber, Moon and Spencer.

5.3.4 Hydrodynamic Analogy

The flow of incompressible fluids and gases without internal friction is very similar to the electric current flow, as a first approximation. (In fact, historically, the laws of electric current flow were patterned after the relations of fluid flow.)

For any fluid, the product of its velocity **v** and its mass density ρ gives its flow density **F** through unit area, i.e.

$$\mathbf{F} = \rho \mathbf{v} \tag{5.14}$$

If the fluid is incompressible and has no sources or sinks, then the total flux through such a closed surface must be zero, i.e. the divergence of its flow density must vanish. Thus,

$$\text{div } \mathbf{F} = \nabla \cdot \mathbf{F} = 0 \tag{5.15}$$

In Cartesian coordinates, this becomes

$$\frac{\partial}{\partial x}(\rho v_x) + \frac{\partial}{\partial y}(\rho v_y) + \frac{\partial}{\partial y}(\rho v_z) = 0 \tag{5.16}$$

This is the equation of continuity.

If the fluid is homogeneous, so that ρ is everywhere the same, then from Eq. (5.15) we get

$$\text{div } \mathbf{v} = \nabla \cdot \mathbf{v} = 0 \tag{5.17}$$

For Irrotational flow of the fluid,

$$\text{curl } \mathbf{v} = \nabla \times \mathbf{v} = 0 \tag{5.18}$$

Hence introducing a scalar potential function, i.e. the velocity potential ϕ (which is +ve by convention),

$$\mathbf{v} = + \text{ grad } \phi = + \nabla \phi \tag{5.19}$$

The equations (5.15) and (5.19) together give

$$\nabla^2 \phi = 0 \tag{5.20}$$

which is the Laplace's equation, which is the same as the operating equation for the electrostatic and the magnetostatic fields. The boundary conditions for the fluid-flows guided by the solid guides are quite analogous to the electric current flow. The lines to which the velocity vector is tangential are the stream-lines and together with the equipotential surface, they form an orthogonal system of gradient lines. When the sources and the sinks are included in the field region, the divergence of **v** does not vanish and the operating equation for the potential becomes

$$\nabla^2 \phi = + \frac{e}{\rho} \tag{5.21}$$

where e is the volume density of the flux. This equation is thus the same as in the case of electrostatics. When the rotational motion of the incompressible fluid is included, then the line integral of **v** enclosing the axis of rotation is no longer zero. As we shall see later, the general solutions of fluid dynamics consist of a superposition of the electrostatic and the magnetostatic field solutions.

Since this technique has been more extensively used for solving the magnetic field problems, we shall discuss it further in the relevant section in the book.

5.4 NUMERICAL METHODS

With the development and availability of high speeds and extremely high storage capability of modern digital computers, the partial differential equations like Laplace, Poisson and others can be solved numerically with relative ease; and these numerical methods of solving the P.D.E.'s are becoming of ever-increasing importance in engineering as well as physics. We shall discuss in some detail the numerical solutions of the Laplace and some allied equations, mainly for the two-dimensional problems though these methods are applicable to a wide range of three-dimensional problems as well. There now exist more than one numerical method for solving these differential equations. In the present discussion we shall restrict ourselves to two basic methods only and follow the historical sequence of development. The two basic methods which we shall describe are:

(a) Finite Difference Methods (F.D.M.)
(b) Finite Element Method (F.E.M.)

5.5 FINITE DIFFERENCE METHODS

Finite difference methods, which can be traced back to Gauss (1823), have been used more generally since about 1940, and can be used for the numerical solutions, of any required accuracy, of both Laplacian and Poissonian fields. In fact, these methods are also applicable to the three-dimensional as well as the time-varying field problems. Even though some of the theoretical considerations associated with the development of these methods are quite difficult, the range of problems which can be solved by these techniques is very wide, such as problems with complex boundary shapes, different boundary conditions, a number of interconnected regions, and take account of physical nonlinearities like the presence of saturating iron, and so on. However there is a major drawback of the F.D.M., which of course is a common constraint of all the numerical methods, and it is that the complete process of the solution has to be carried out for each set of parameters defining any problem.

The solutions produced by the F.D.M. are in the form of values (of the function describing the field) at discrete points which are prearranged in some regular manner over the whole region of the field. These values are obtained by replacing the original partial differential equation by a large number of finite difference equations taking the form of linear algebraic equations which connect the potentials at each point with the potentials at the points in the immediate vicinity. Thus the problem is finally reduced to the solution of a set of simple simultaneous algebraic equations for the potential values. Since a very large number of such equations normally occur in a problem, it is quite inconvenient to solve them by direct elimination or matrix inversion. Since computers are now available, some sort of iterative technique which exploits the simple form of these equations, is used. When such computations were initially done by hand, the relaxation method (developed by Southwell) was extensively used. While the hand-relaxation is very flexible, the effectiveness of the method for quick solution is dependent on the experience and the skill of the person concerned. But when the problem is solved on the machine, it is tackled by an entirely fixed cycle of operation, (each cycle being called iteration). These methods are designed to take advantage of the digital computing techniques, and in optimum forms give remarkably rapid solutions. Since 1950s, a significant number of different methods have been developed and the relevant literature is very extensive. In this chapter, we have kept the treatment of the methods quite simple and at an introductory level. Readers interested in the depth of these

methods, should study the original papers, some of which are listed in the bibliography and the books by Forsythe & Wasow, Todd, Varga, Ames, Smith, Wachspress, Walsh, and Mitchel. We shall now establish the basic F.D. equations and discuss the various iteration techniques as well as some of their applications to the different types of problems. For generality, the equations are developed in terms of the Poisson's equation, since in the Laplace's equation the source term gets reduced to zero.

5.5.1 Finite Difference Representation

The first step in reducing the field P.D.E. to a set of F.D. equations is to select a suitable spatial distribution of points in the region. Since the problem is being solved on the computer, it is obvious that a regular distribution of points in the region be chosen so that the same form of the F.D. equation is satisfied at all the points and the problem representation gets simplified. Such distributions are given by the arrangements of points lying at the 'nodes' of any regular network or 'mesh'. There are three different regular mesh systems available as shown in Figure 5.7.

1. The square mesh (also known as symmetrical star or molecule)
2. The equilateral triangular mesh
3. The equiangular hexagonal mesh

There are also two irregular meshes which are very useful (see Figure 5.7).

1. The polar mesh
2. The rectangular mesh

Here we shall restrict our discussion to the square meshes and the rectangular meshes, the latter being used in conjunction with the square meshes when the number of nodes in a problem

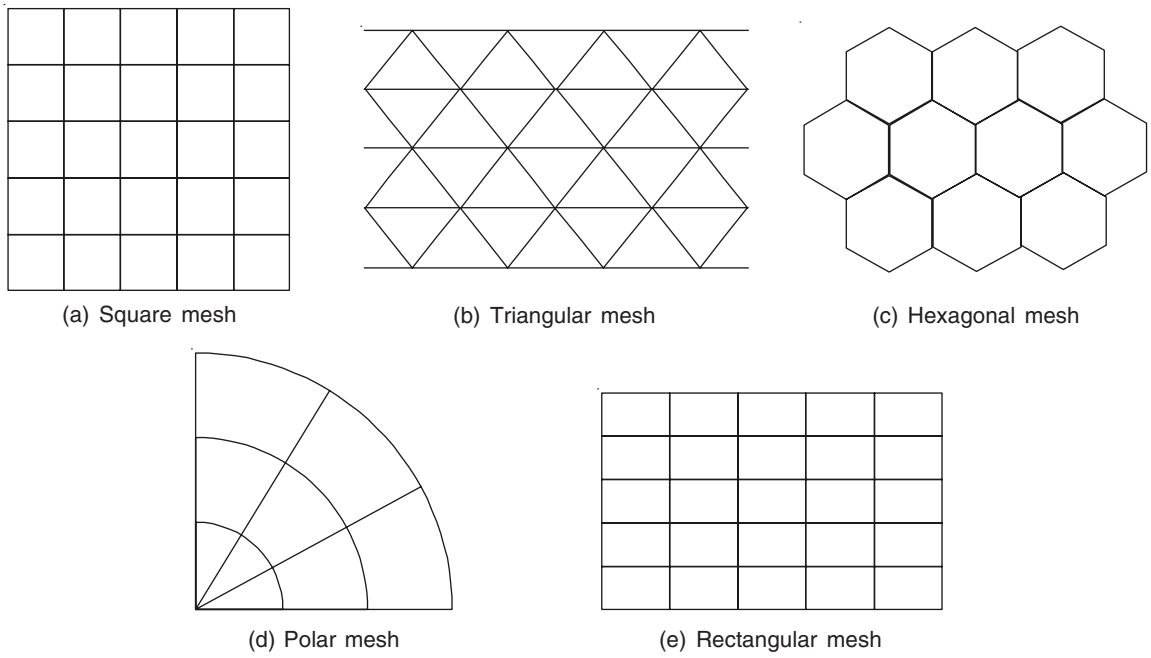
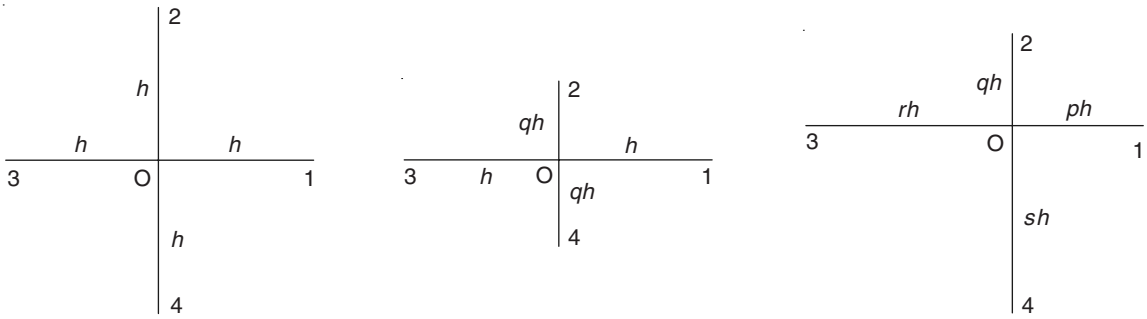


Figure 5.7 Different types of meshes.

have to be minimized. For the equations and the other details of these meshes, the reader should consult the books by Allen, Forsythe & Wasow, Smith, Boast, Richtmayer, and Southwell.

5.5.2 Basic Equations for the Square and the Rectangular Meshes

We shall consider the symmetrical star of Figure 5.8(a), which is relevant at all the nodes in a uniform square mesh. However adjacent to those boundaries where the mesh and the boundaries do not coincide, and in the regions where the meshes of different sizes meet, the arms of the centre node O will be of different lengths. This qualification applies to the rectangular meshes as well. So, for the sake of generality of the problem, we shall consider an asymmetrical star of Figure 5.8(c). In this asymmetrical star, let the distances of the points 1, 2, 3, 4 from O be *ph*, *qh*,



(a) Symmetrical star (b) Rectangular star (c) Asymmetrical star

Figure 5.8 Symmetrical and asymmetrical stars.

rh, and *sh*, respectively, where *p*, *q*, *r*, *s* are less than unity, and *h* is the distance between the lines in the regular portion of the mesh. It is usual to call *h* as the *mesh length*. It is usually equal to the side of a square of the square mesh, and also the longer of the two arms of the rectangular mesh.

Next we consider the reduction of the operating P.D.E. to the relevant difference equation. There are three different ways in which this can be done. They are:

1. Taylor series expansion
2. Integration method
3. Rayleigh–Ritz method or Variational method.

The first method is the most commonly used one, though the second one gives the best results, and the third method shows the equivalence between the F.D.M. and the F.E.M. For the details of these methods, the reader is recommended to study Varga, and Browne.

By the Taylor series method, the derivative terms, i.e. $\partial^2\phi/\partial x^2$ and $\partial^2\phi/\partial y^2$ of the Poisson's equation are reduced to the difference form. At any point x on the line parallel to the x-axis, passing through the point O, the potential ϕ can be expanded in terms of the value of the potential ϕ_0 at O by using the Taylor series. This is done for the points 1 and 3, respectively, as

$$\phi_1 = \phi_0 + ph\left(\frac{\partial\phi}{\partial x}\right)_0 + \left(\frac{1}{2!}\right)p^2h^2\left(\frac{\partial^2\phi}{\partial x^2}\right)_0 + \left(\frac{1}{3!}\right)p^3h^3\left(\frac{\partial^3\phi}{\partial x^3}\right)_0 + \left(\frac{1}{4!}\right)p^4h^4\left(\frac{\partial^4\phi}{\partial x^4}\right)_0 + \cdots \quad (5.22)$$

$$\phi_3 = \phi_0 - rh\left(\frac{\partial\phi}{\partial x}\right)_0 + \left(\frac{1}{2!}\right)r^2h^2\left(\frac{\partial^2\phi}{\partial x^2}\right)_0 - \left(\frac{1}{3!}\right)r^3h^3\left(\frac{\partial^3\phi}{\partial x^3}\right)_0 + \left(\frac{1}{4!}\right)r^4h^4\left(\frac{\partial^4\phi}{\partial x^4}\right)_0 - \cdots \quad (5.23)$$

Adding r times the Eq. (5.22) with p times the Eq. (5.23), we get

$$r\phi_1 + p\phi_3 = (p+r)\phi_0 + \left(\frac{h^2}{2!}\right)pr(p+r)\left(\frac{\partial^2\phi}{\partial x^2}\right)_0 + \left(\frac{h^3}{3!}\right)pr(p^2-r^2)\left(\frac{\partial^3\phi}{\partial x^3}\right)_0 + \left(\frac{h^4}{4!}\right)pr(p^3+r^3)\left(\frac{\partial^4\phi}{\partial x^4}\right)_0 + \cdots \quad (5.24)$$

Ignoring the terms containing the higher powers of h from h^3 onwards, we get

$$h^2\left(\frac{\partial^2\phi}{\partial x^2}\right)_0 = \frac{2\phi_1}{p(p+r)} + \frac{2\phi_3}{r(p+r)} - \frac{2\phi_0}{pr} \quad (5.25)$$

In a similar manner, the expression for $\partial^2\phi/\partial y^2$ is obtained as

$$h^2\left(\frac{\partial^2\phi}{\partial y^2}\right)_0 = \frac{2\phi_2}{q(q+s)} + \frac{2\phi_4}{s(q+s)} - \frac{2\phi_0}{qs} \quad (5.26)$$

Hence substituting in the Poisson's equation, we get

$$2\left[\frac{\phi_1}{p(p+r)} + \frac{\phi_2}{q(q+s)} + \frac{\phi_3}{r(p+r)} + \frac{\phi_4}{s(q+s)} - \left\{\frac{1}{pr} + \frac{1}{qs}\right\}\phi_0\right] + h^2W = 0 \quad (5.27)$$

where W is the right-hand term of the Poisson's equation.

This general form of the difference equation is not often required, but from it by substituting the appropriate values of p, q, r, s, any necessary equation can be derived immediately. For a uniform square mesh as in Figure 5.8(a), the difference equation becomes, when $p = q = r = s = 1$,

$$\phi_1 + \phi_2 + \phi_3 + \phi_4 - 4\phi_0 + h^2W = 0 \quad (5.28)$$

For a rectangular mesh of Figure 5.8(b), $p = r = 1$ and $q = s$ in Eq. (5.27) gives

$$\phi_1 + \frac{\phi_2}{q^2} + \phi_3 + \frac{\phi_4}{q^2} - 2\left(1 + \frac{1}{q^2}\right)\phi_0 + h^2W = 0 \quad (5.29)$$

For the Laplace's equation, $W = 0$ and so we get

$$\phi_1 + \phi_2 + \phi_3 + \phi_4 - 4\phi_0 = 0 \quad (5.30)$$

It should be remembered that the above difference equations are approximations to the field equations, and the higher order terms of Eq. (5.24) have been neglected. This error is called the 'truncation error', and has to be investigated carefully though in general terms it can be said that h can be so chosen for a problem that the error can be made smaller and smaller.

We shall next derive the equivalent difference equations by the integration method, and consider the same asymmetrical 5-point star of Figure 5.8(c), which is repeated in Figure 5.9. This method was first introduced by McNeal in 1953 in order to obtain a symmetrical coefficient matrix for the finite difference equations when an irregular mesh and curved boundaries are encountered. The key idea of the method is to integrate the field equation over the mesh regions so chosen that they have an averaging effect which results in the symmetry of the matrix. As usual the domain of the solution is covered with the mesh, at the nodes of which the values of the field function are sought. The mesh need not be rectangular but must be so chosen that any internal

interfaces must coincide with the lines of the mesh. At the mid-point of each mesh-link (or mesh-arm) connecting the adjacent nodes, perpendicular lines are drawn which form a polygonal region surrounding the node. In our present problem, this is the rectangular region ABCDA surrounding the node O as shown in Figure 5.9. We then integrate the field equation at each node over the

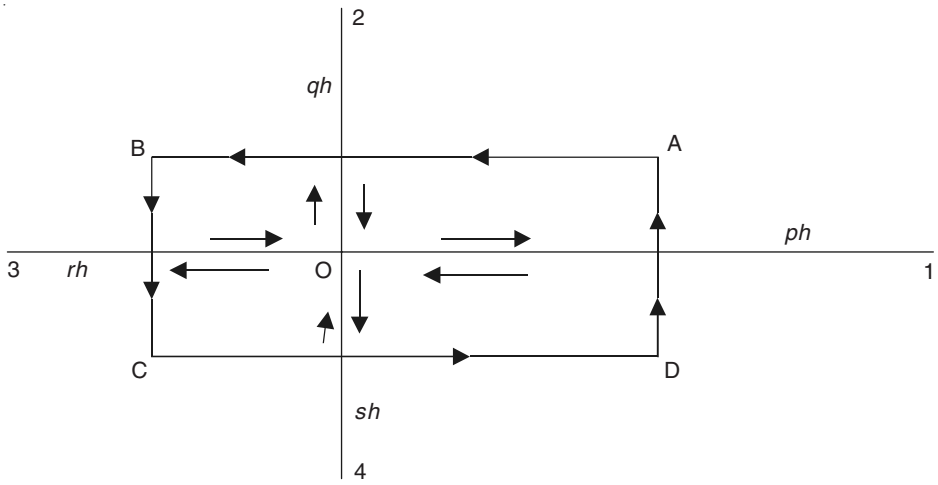


Figure 5.9 Asymmetrical 5-point star with the integration paths shown.

associated polygonal area, i.e. the rectangular region ABCDA in this case. Then by using the Green's theorem, we transform the surface integrals to line integrals and then the first-order derivative terms are approximated by the central difference formulae, thereby arriving at the finite difference equation at each node.

We start with the Poisson's equation over the region of the node under consideration,

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + W = 0$$

Integrating this equation over the surface enclosed by the rectangular path ABCDA, we get

$$\iint_{S\text{ABCDA}} \left\{ \left(\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} \right) dx\, dy \right\} + \iint_{S\text{ABCDA}} W\, dx\, dy = 0 \tag{5.31}$$

Applying the Green's theorem to this equation, it gets expressed in terms of the line integrals taken in the counterclockwise direction:

$$\oint_{\text{ABCDA}} \left\{ \left(\frac{\partial \phi}{\partial x} \right) dy - \left(\frac{\partial \phi}{\partial y} \right) dx \right\} + \iint_{S\text{ABCDA}} W\, dx\, dy = 0 \tag{5.32}$$

Adding the expressions for each sub-region (as shown in Figure 5.9), and then approximating the first-order derivative terms by the central difference formulae, the following finite difference equation is obtained.

$$\left[\left(\frac{\phi_1 - \phi_0}{ph} \right) \left(\frac{qh + sh}{2} \right) - \left(\frac{\phi_0 - \phi_3}{rh} \right) \left(\frac{qh + sh}{2} \right) + \left(\frac{\phi_2 - \phi_0}{qh} \right) \left(\frac{ph + rh}{2} \right) - \left(\frac{\phi_0 - \phi_4}{sh} \right) \left(\frac{ph + rh}{2} \right) \right] + W \left[(p + r)(q + s) \frac{h^2}{4} \right] = 0$$

or

$$\phi_1\left(\frac{q+s}{2p}\right)+\phi_2\left(\frac{p+r}{2q}\right)+\phi_3\left(\frac{q+s}{2r}\right)+\phi_4\left(\frac{p+r}{2s}\right)-\phi_0\left[\left(\frac{q+s}{2p}\right)+\left(\frac{p+r}{2q}\right)+\left(\frac{q+s}{2r}\right)+\left(\frac{p+r}{2s}\right)\right]+\left[\frac{(p+r)(q+s)h^2}{4}\right]W=0 \tag{5.33}$$

If in this equation we make $p = q = r = s = 1$, then the above equation becomes

$$\phi_1 + \phi_2 + \phi_3 + \phi_4 - 4\phi_0 + h^2W = 0$$

which is the same as that for the square mesh in the Poisson's equation [see Eq. (5.28)]. By making $W = 0$, further, it reduces to the difference equation for the Laplace's equation as in Eq. (5.30). For the rectangular mesh, by substituting $p = r = 1$ and $q = s$, we get from Eq. (5.33),

$$\phi_1 q + \frac{\phi_2}{q} + \phi_3 q + \frac{\phi_4}{q} - \phi_0\left(2q + \frac{2}{q}\right) + qh^2W = 0$$

which is the same as Eq. (5.29).

For further details, generalizations and applications to other problems, such as for regions with eddy current distributions and with different magnetic permeabilities, the reader is referred to Browne.

Thus it is seen that for the internal nodes, both the methods, i.e. the Taylor series expansion and the integration method, give identical difference equations. The advantage gained by starting with the integration method is due to the presence of the different types of boundaries and boundary conditions which result in a 'cleaner' and well-behaved coefficient matrix. It is the derivation of the coefficient matrix that we shall discuss next.

5.5.3 Reduction of the Field Problem into a Set of Simultaneous Equations

Let us consider a simple Poissonian field region within the square boundary as shown in Figure 5.10.

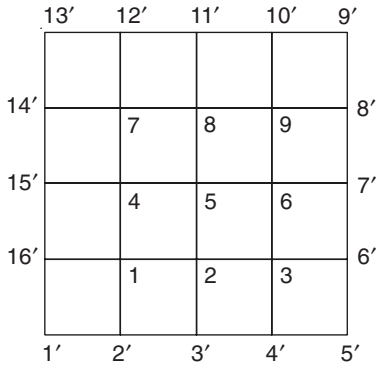


Figure 5.10 Poissonian field region.

Let the nodes on the boundary be denoted by the primed numbers 1′ to 16′ and the internal nodes be numbered 1 to 9. The value of W and the values of the potentials on the boundary ϕ_1' to

ϕ'_{16} are given and the values of the potentials ϕ_1 to ϕ_9 at the interior nodes have to be determined. For this purpose we use Eq. (5.28) at each of the interior nodes. Hence at the node 1, we get

$$\phi_2 + \phi_4 + \phi'_{16} + \phi'_2 - 4\phi_1 + h^2W = 0$$

and at the node 2, we have

$$\phi_3 + \phi_5 + \phi_1 + \phi'_3 - 4\phi_2 + h^2W = 0.$$

We derive similar equations for the nodes 3 to 9. The nine potentials ϕ_1 to ϕ_9 are thus related amongst themselves and to the boundary conditions by nine linear algebraic equations, and they (the potentials) can be evaluated by solving these equations simultaneously. In a similar way, any field problem can be solved as a set of simultaneous linear difference equations, whatever be the shapes and the nature of the boundaries. We shall now consider the set of these simultaneous equations in the matrix form. If **C** is the matrix of the coefficients of the unknown potentials, **u** is the column matrix of the unknown potentials ϕ_n (in this case ϕ_1 to ϕ_9), and **v** is the column matrix of the sum of the known boundary potentials and the constant terms are h^2W (when and wherever present), the equations can be expressed in a compact matrix form as

$$\mathbf{Cu} = \mathbf{v} \tag{5.34}$$

The small number of (nonzero) terms in each difference equation results in a correspondingly small number in each row of **C** and this is the basic feature of the matrix which is exploited in an iterative solution. It is also obvious that the elements of **C** depend on the particular type of the difference equation used and not whether the field is Laplacian or Poissonian, and that **C** is essentially diagonally dominant and symmetrical about the leading diagonal. Thus the nature of any iterative scheme applied to solve Eq. (5.34) is decided upon by the properties of the coefficient matrix **C**. We will describe only some of the basic points of the methods used, and for the detailed information regarding these methods, the reader is recommended to study the books by Varga, Wachspress, Lawrenson & Binns, Ames, Smith, and Mitchell.

5.5.4 Computational Methods

The hand computation was first used by Gauss, but not until it was rediscovered and developed by Southwell and his coworkers in 1930–40 that its flexibility and power appreciated by the engineers and the physicists. (The interested readers are recommended the books by Southwell, Allen, and Shaw.) Since the late fifties, this method has been displaced by the machine iteration which has permitted the treatment of much larger and more varied problems. But the method (Relaxation) is still useful for smaller problems and those with complex boundary shapes. Furthermore, since the understanding of this method helps the later discussion of the machine methods, we will start with a brief explanation of the essential points of the method. The relaxation method depends on the concept of the 'residual'. When the potentials ϕ_0, ϕ_1, ϕ_2, ϕ_3, ϕ_4 are so chosen as to satisfy a difference equation, say Eq. (5.28), its right-hand side is zero. But when the potentials do not satisfy this equation, then the right-hand side is not equal to zero but to some other quantity R_0 which is given by

$$R_0 = \phi_1 + \phi_2 + \phi_3 + \phi_4 - 4\phi_0 + h^2W \tag{5.35}$$

where R_0 is called the *residual* of the difference equation at the point O, and the relaxation method aims at systematic reduction of R_0 to zero for all the equations. When the residuals of all the difference equations are zero, the values of the potentials correspond to an exact solution. But,

in practice, it is sufficient to reduce the residuals to low values, to obtain a good approximation to the exact solution. In fact, there is no exact relationship between the magnitudes of the residuals and the accuracy of the potentials. But generally, satisfactory solutions can be obtained when the following conditions are satisfied:

1. Individual values of the residuals should be reduced to about 0.1% of the mean value of the potentials.
2. The algebraic sum of all the residuals should be of magnitude similar to that of the individual residuals.
3. The residuals, with respect to their magnitudes and signs, should be uniformly mixed over the whole field region.

The technique of minimizing the residuals is called the 'Point Relaxation'. At the node under consideration (at start assumed to be the centre of a symmetrical star) the value of the residual is calculated by using Eq. (5.35) and then this residual is reduced to zero by a change in ϕ_0 of $R_0/4$. This change in ϕ_0 produces an equal change in the residuals at each of the adjacent nodes in the star, and applying the same equation to each of these nodes in turn, they experience a change of $R_0/4$ which is same as the change in ϕ_0 of the original central node. Thus a residual of $4R$ at the central node of a star can be reduced to zero by an increase in the potential of the magnitude R at that node and by the addition of R to the residuals at each of the points of the star. This process is called 'spreading' the residual, and is explained pictorially in Figure 5.11. It

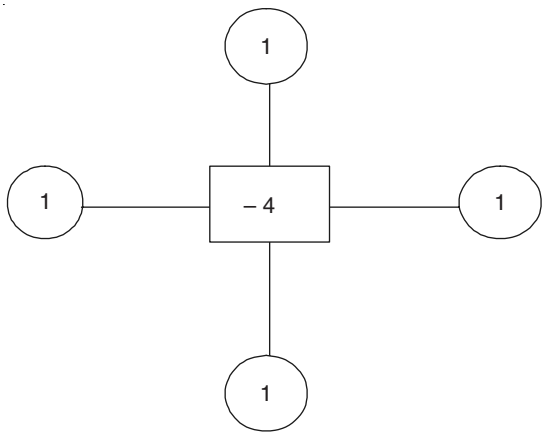


Figure 5.11 Relaxation pattern.

must be remembered that in the process of spreading of the residuals, the boundary points are not affected. Similar patterns can be worked out for asymmetrical stars. The rate of convergence of the solution can be increased significantly by the following two important means.

(a) The degree of relaxation applied to a node
(b) Relaxing several points simultaneously

Over- and under-relaxation. When the potential at a point is changed by an amount which reverses the sign of the residual, the operation is called *over-relaxation*. When the change in the residual is insufficient to reduce it to zero, then it is called *under-relaxation*. Both these operations can greatly increase the convergence of the solution.

Line- and block-relaxation. The process of convergence gets accelerated if instead of relaxing each star individually, we relax a group of points, either in a line or in a block, simultaneously. This is done by using a line- or block-relaxation operator, the pattern for which is obtained by a simple addition of point patterns suitably oriented as required by the arrangement. This makes the process both economical as well as faster.

With this background information about the hand-relaxation method, we now consider the methods of machine computation, i.e. iteration. The technique on the machine is still the same, i.e. modify continually the values of the potentials till all the equations are satisfied to a sufficient degree of accuracy. Considerations of stability, accuracy, etc. for the two methods are similar, but there are some important features which distinguish the machine methods from the hand methods. Firstly, the machine methods are designed for use in a completely automatic cycle in which each equation is considered in a fixed sequence throughout the calculation. Secondly, the improved values of the potentials are determined directly from the difference equations, and hence the concept of the residual is disregarded. The implications are that the rate of convergence of a basic iterative method is slower than that of the relaxation method in the hands of an expert, and also since the method can be easily programmed for the machine, the solutions can be obtained very speedily. So now we consider the basic concepts of the iterative schemes, which are the ordering of the cycles and the examination of the convergence.

Designation of the nodes. As mentioned earlier, each node is considered in turn in a fixed cycle, and so each node is represented by a number pair. We consider a basic rectangular array of $(p + 1) \times (q + 1)$ points on which the boundary is superposed (Figure 5.12). Any node can be

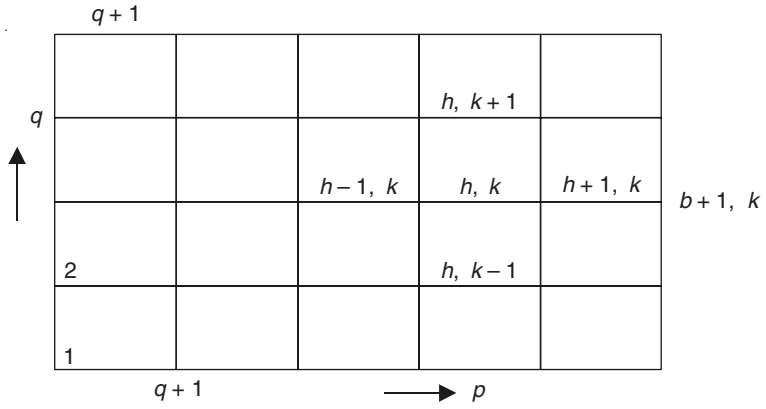


Figure 5.12 A rectangular array of points.

specified by a number pair (h, k), where $1 \le h \le (p + 1)$ and $1 \le k \le (q + 1)$, or by a single number $(h - 1)(q + 1) + (k + 1)$ where 1 is taken as the point at the lower left-hand corner, and the columns of the points are scanned in each cycle from the left to the right (increasing p) and from the bottom to the top (increasing q).

Examination of convergence. In the iterative process, since all he relevant information is stored inside the computer, it is difficult to observe the distribution of the residuals, and hence this guide to the convergence cannot be easily used. Since in the machine, the computation is much easier, the permissible upper limit of the residual can be set to a much lower value, say, 0.0001% compared with the manual relaxation, and also check upon the residual after, say, 10 or 20

iterations at a time, in order to minimize the extra work to be done by the machine for the control. Alternatively, we can disregard the residual completely and to test for the convergence, we study the change in the potential, which is called the 'displacement' occurring at each node, and the iteration procedure is stopped when the changes are sufficiently small, say, 10^{-6} % of the potential values. This is a good test, though sometimes it can be argued that the small changes can mean slow convergence rather than high accuracy. We shall discuss this point later. We now discuss the basic elements of some of the iterative procedures.

Jacobi method (*or Richardson method or the method of simultaneous displacement*). This is the simplest of the iterative methods. Each new value of the potential at the centre of a star is determined as that which satisfies the basic difference equation in terms of the previous values of the potentials at the other points. So if ϕ^n represents the potential after the nth iterative cycle, then the Jacobi form of the symmetrical Poisson difference equation is

$$\phi_0^{n+1} = \left(\frac{1}{4}\right)\left(\phi_1^n + \phi_2^n + \phi_3^n + \phi_4^n + h^2 W\right) \tag{5.36}$$

Or, in more general terms, using the double-letter notation defined earlier, we have

$$\phi_{h,k}^{n+1} = \left(\frac{1}{4}\right)\left(\phi_{h+1,k}^n + \phi_{h,(k+1)}^n + \phi_{h-1,k}^n + \phi_{h,(k-1)}^n + h^2 W\right) \tag{5.37}$$

Similar equations can be derived for the asymmetrical patterns. The convergence of the Jacobi method is poor. Furthermore, two sets of the potential values [for the nth and the $(n + 1)$th iterations] have to be stored in the memory. Because of these disadvantages, this method is not used in practice.

Gauss-Seidel method (*or Liebmann method or the method of successive displacement*). This is a simple modification of the previous method in which the most recently computed values of the potentials are used. If the nodes are scanned column by column, from the left to the right and starting at the bottom of each, the general form of the symmetrical Poisson difference equation is

$$\phi_{h,k}^{n+1} = \left(\frac{1}{4}\right)\left(\phi_{h+1,k}^n + \phi_{h,(k+1)}^n + \phi_{h-1,k}^{n+1} + \phi_{h,(k-1)}^{n+1} + h^2 W\right) \tag{5.38}$$

This method is economical because it requires the storage of only one complete set of potentials, and also converges twice as quickly as the Jacobi method, and yet it is uneconomically slow compared with the more sophisticated methods.

Successive over-relaxation method (*S.O.R. method*). This method is also called the Extrapolated Liebmann method and was discovered independently by Frankel and Young. It is a very flexible method and also rapidly convergent. It can be derived from the Gauss-Seidel method by introducing a factor α, and the new value of the potential is obtained as the sum of the old value and α times the difference between the value given by Eq. (5.38), and the old value of ϕ, i.e.

$$\phi_{h,k}^{n+1} = \phi_{h,k}^n + \left(\frac{\alpha}{4}\right)\left(\phi_{h+1,k}^n + \phi_{h,(k+1)}^n + \phi_{h-1,k}^{n+1} + \phi_{h,(k-1)}^{n+1} + h^2 W - 4\phi_{h,k}^n\right) \tag{5.39}$$

where α is a convergence or a relaxation factor which determines the degree of over-relaxation, and it can be shown that it must be between 1 and 2. When $\alpha = 1$, Eq. (5.39) reduces to Eq. (5.38) for the Gauss-Seidel method, and when $\alpha \geq 2$, the process becomes unstable. When α lies between

the specified limits, the convergence rate is higher than that for $\alpha = 1$, and for some optimum value α_b which is different for every problem, the rate is greatly increased.

The convergence of the solution. Varga has shown theoretically that for a stable convergent process, the number of iterations N required to reduce the largest error at any node to a fraction ε of some previous value, is given by the relation

$$N \doteqdot - F \log \varepsilon \tag{5.40}$$

where F, the asymptotic rate of convergence is a function of the shape and the condition of the boundary, the particular type of the difference equation, and the convergence factor. It is defined by

$$F = - \log \mu \tag{5.41}$$

where μ can be obtained as the limiting value of the ratio of the absolute values of the maximum changes in the potentials occurring on successive iterations when the convergence factor is unity (Gauss-Seidel). Hence, if

$$U^n = \max \left| \phi_m^n - \phi_m^{n+1} \right| \tag{5.42}$$

then

$$\mu = \lim_{n \to \infty} \frac{U^{n+1}}{U^n}, \tag{5.43}$$

where μ is also the largest eigenvalue or the spectral radius of the iteration matrix characterizing the S.O.R. scheme. So we now write down the matrix Eq. (5.34) for the set of the difference equations of the S.O.R. in the form

$$(\mathbf{I} - \mathbf{L} - \mathbf{U})\mathbf{u} = \mathbf{v} \tag{5.44}$$

where $\mathbf{I}$ is the unit matrix of the order m, and $\mathbf{L}$ and $\mathbf{U}$ are the lower and the upper triangular matrices with null diagonals. Hence

$$\mathbf{I}\,\mathbf{u} = \mathbf{u} = (\mathbf{L} + \mathbf{U})\mathbf{u} + \mathbf{v} \tag{5.45}$$

where

$$- \mathbf{L} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & . & 0 \\ a_{21} & 0 & 0 & 0 & 0 & 0 & . & 0 \\ a_{31} & a_{32} & 0 & 0 & 0 & 0 & . & 0 \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ a_{m1} & a_{m2} & . & . & . & . & a_{m\,m-1} & 0 \end{bmatrix}$$

and

$$- \mathbf{U} = \begin{bmatrix} 0 & a_{12} & a_{13} & a_{14} & . & . & . & a_{1m} \\ 0 & 0 & a_{23} & a_{24} & . & . & . & a_{2m} \\ 0 & 0 & 0 & a_{34} & . & . & . & a_{3m} \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{m-1\,m} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The Jacobi iteration scheme is then

$$\mathbf{u}^{n+1} = [\mathbf{L} + \mathbf{U}]\mathbf{u}^n + \mathbf{v} \tag{5.46}$$

and the Gauss-Seidel scheme is

$$\mathbf{u}^{n+1} = \mathbf{L}\mathbf{u}^{n+1} + \mathbf{U}\mathbf{u}^n + \mathbf{v} \tag{5.47}$$

This reduces to

$$\mathbf{u}^{n+1} = [\mathbf{I} - \mathbf{L}]^{-1}\,\mathbf{U}\mathbf{u}^n + [\mathbf{I} - \mathbf{L}]^{-1}\,\mathbf{v}$$

Note: The matrices $[\mathbf{L} + \mathbf{U}]$ and $[\mathbf{I} - \mathbf{L}]^{-1}\,\mathbf{U}$ are called the Jacobi and the Gauss-Seidel matrix, respectively, of the coefficient matrix **C**.

And the S.O.R. scheme gets defined by

$$\mathbf{u}^{n+1} = \mathbf{u}^n + \alpha\,[\mathbf{L}\mathbf{u}^{n+1} + \mathbf{U}\mathbf{u}^n + \mathbf{v} - \mathbf{u}^n] \tag{5.48}$$

or

$$\mathbf{u}^{n+1} = \{(\mathbf{I} - \alpha\,\mathbf{L})^{-1}\,[\alpha\,\mathbf{U} - (\alpha - 1)\,\mathbf{I}]\}\mathbf{u}^n + \alpha\,(\mathbf{I} - \alpha\,\mathbf{L})^{-1}\,\mathbf{v} \tag{5.49}$$

In this equation, the term in the curly brackets is the iteration matrix. The error matrix at any stage is the difference between the true and the approximate solutions, i.e.

$$\mathbf{e}^n = \mathbf{u} - \mathbf{u}^n \tag{5.50}$$

Also the amount by which the ϕ's or the **u** change during one iteration, i.e. $\mathbf{u}^{n+1} - \mathbf{u}^n$ is called the 'displacement vector' **d**, so that

$$\mathbf{u}^{n+1} = \mathbf{u}^n + \mathbf{d} \tag{5.51}$$

Equation (5.49) gives the potential values after $(n + 1)$ iterations. To find the error, let us look at the equation for the exact solution, i.e.

$$\mathbf{u} = \{(\mathbf{I} - \alpha\,\mathbf{L})^{-1}\,[\alpha\,\mathbf{U} - (\alpha - 1)\mathbf{I}]\}\mathbf{u} + \alpha\,(\mathbf{I} - \alpha\,\mathbf{L})^{-1}\,\mathbf{v} \tag{5.52}$$

From these three Eqs. (5.49), (5.50) and (5.52), the initial error $\mathbf{e}^0$ is propagated into the successive iterations by the equation

$$\mathbf{e}^{n+1} = \{(\mathbf{I} - \alpha\,\mathbf{L})^{-1}\,[\alpha\,\mathbf{U} - (\alpha - 1)\,\mathbf{I}]\}^{n+1}\,\mathbf{e}^0 \tag{5.53}$$

and the successive error vectors are related by the equation

$$\mathbf{e}^{n+1} = \mathbf{H}\mathbf{e}^n \tag{5.54}$$

where **H** the error matrix is found to be the same as the iteration matrix defined above. It should be noted that **H** is independent of n. Assuming that **H** has m linearly independent eigenvectors $\mathbf{v}_s$ corresponding to the eigenvalues α_s, the initial error vector $\mathbf{e}^0$ (with its m components) can be expressed as a linear combination of the eigenvectors in the form

$$\mathbf{e}^0 = \sum_{s=1}^{m} \mathbf{c}_s\,\mathbf{v}_s$$

and so the error vector after n iterations is

$$\mathbf{e}^n = \sum_{s=1}^{m} \mathbf{c}_s\,\mu_s^n\,\mathbf{v}_s$$

from which it follows that $\mathbf{e}^n$ tends to zero, as n increases, when the largest value of $|\mu_s|$, (s = 1, 2, 3, ..., m), called the *spectral radius* of **H** is less than unity.

When **u** is not the exact solution of **Cu** = **v,** the difference **v** − **Cu** is called the *residual vector* **r**. In the manual relaxation method, the components of **r** are calculated for an initial estimate of $\mathbf{u}^0$ and then we reduce the residuals to zero by making the appropriate changes in the components of **u**. In the hand relaxation method, the reduction of the residuals need not be done in any systematic order, but in the computer, this order is systematized and done in a sequential manner. Thus, the current value of the residual vector $\mathbf{r}_c$ for the (n + 1)th iterative cycle is

$$\mathbf{r}_c = \mathbf{v} - \mathbf{Cu} = \mathbf{v} - [\mathbf{I} - \mathbf{L} - \mathbf{U}]\mathbf{u} = \mathbf{v} - \mathbf{u}^n + \mathbf{L}\mathbf{u}^{n+1} + \mathbf{U}\mathbf{u}^n \tag{5.55}$$

When $\mathbf{r}_c$ is made zero, we get a vector equation for the next iterative cycle of $\mathbf{u}^n$, i.e. $\mathbf{u}^{n+1}$, i.e.

$$0 = \mathbf{v} - \mathbf{u}^{n+1} + \mathbf{L}\mathbf{u}^{n+1} + \mathbf{U}\mathbf{u}^n$$

$$\therefore \qquad \mathbf{u}^{n+1} = \mathbf{L}\mathbf{u}^{n+1} + \mathbf{U}\mathbf{u}^n + \mathbf{v} \tag{5.56}$$

This equation is the same as Eq. (5.47) of the Gauss-Seidel scheme and hence that is a variation of the systematic relaxation method which reduces the successive elements of the current residual to zero. The current residual vector is thus equal to the displacement vector in the Gauss-Seidel iteration, as Eq. (5.56) can also be written as

$$\mathbf{u}^{n+1} = \mathbf{u}^n + (\mathbf{v} - \mathbf{u}^n + \mathbf{L}\mathbf{u}^{n+1} + \mathbf{U}\mathbf{u}^n) = \mathbf{u}^n + \mathbf{r}_c$$

Thus the rate of convergence of this scheme can be further improved by the S.O.R. method, defined by

$$\mathbf{u}^{n+1} = \mathbf{u}^n + \alpha\mathbf{r}_c$$

where α the relaxation factor is a real positive constant which over-relaxes for $\alpha > 1$ and under-relaxes for $\alpha < 1$. It has been proved that the S.O.R. method converges for $0 < \alpha < 2$ whenever **C** is a symmetric positive definite matrix which is frequently the case with the F.D. equations associated with these P.D.Es.

The optimum value α_b of α for the maximum rate of convergence is given by

$$\alpha_b = \frac{2}{1 + \sqrt{1 - \mu^2}} \tag{5.57}$$

where μ is the spectral radius of the Jacobi iteration matrix (**L** + **U**) associated with the matrix **C** subject to the 'Property A'. This is the value of α that minimizes the spectral radius of the S.O.R. iteration matrix $\{(\mathbf{I} - \alpha\,\mathbf{L})^{-1}\,[\alpha\,\mathbf{U} - (\alpha - 1)\,\mathbf{I}]\}$ when considered as a function of α. The minimum spectral radius is (α_b − 1) and for the convergence this must be < 1, and Eq. (5.57) shows that this condition is fulfilled when $\mu < 1$. As the spectral radius of the corresponding Gauss-Seidel iteration matrix $\{\mathbf{I} - \mathbf{L}\}^{-1}\,\mathbf{U}$ is μ^2, it follows that when the Jacobi iteration matrix converges, both the Gauss-Seidel and the optimized S.O.R. methods must converge.

'Property A' and 'Consistent Ordering'. The relationships between the Jacobi, the Gauss-Seidel, and the S.O.R. iterative methods for solving **Cu** = **v** have been established only for matrices possessing what is called the 'Property A' and for a scanning of the mesh points in what is called a 'Consistent Order'.

'Property A', invented by Young, states that a matrix **C** = [**I** − **L** − **U**] can be transformed by a sequence of permutations into a 'diagonally block-tridiagonal form'

$$\begin{bmatrix} I_1 & U_1 & 0 & . & . & . & . \\ L_2 & I_2 & U_2 & . & . & . & . \\ 0 & L_3 & I_3 & U_3 & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & L_r & I_r \end{bmatrix} \tag{5.58}$$

where I_i are the diagonal matrices of various orders, but made up of the elements of **I**; and L_i are the rectangular matrices made up of the elements of **L;** and U_i are the rectangular matrices made up of the elements of **U**. It should be noted that a sequence of permutations does nothing more than reorder the rows and the columns in the same way. The important thing is that the permuted matrix has the same eigenvalues as the original matrix.

To illustrate the point, let us assume that we are solving a P.D.E. in two independent variables x and y over a closed area S, and the derivative $\dfrac{\partial^2 \phi}{\partial x \partial y}$ is not present. Subdivide S into rectangular meshes and label the mesh points black (B) and white (W) in a chess-board fashion like below

B	W	B	W	B	W	.	.	.
W	B	W	B	W	B	.	.	.
B	W	B	W	B	W	.	.	.
W	B	W	B	W	B			
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.

If now the finite difference equation approximating the differential equation at the point B can be solved for the function value at B exclusively in terms of the function values at the point W, and vice-versa, then the matrix **C** of the difference equation has the 'Property A'.

Or, more formally, sub-divide the positive integers into two classes B and W. Then the matrix **C** has the 'Property A', if the suffices i and j of the nonzero off-diagonal elements c_{ij} are such that if i is placed in B and j in W or vice versa, then no member of B is a member of W. It can also be shown that if the matrix **C** has the 'Property A', then it is always possible to reorder the equations and the unknowns so that the matrix of the reordered set has either the tri-diagonal partitioned form (shown before) or the partitioned form.

$$\begin{bmatrix} D_1 & F \\ E & D_2 \end{bmatrix} \tag{5.59}$$

where the Ds are the square diagonal sub-matrices (not necessarily of the same order).

Consistent ordering. In the Jacobi method , the rate of convergence is independent of the order in which the mesh points are scanned. But the rates of convergence of the Gauss-Seidel and the S.O.R. methods depend on the way the mesh points are scanned because the $(n + 1)$th iterates, as soon as they are available, are required to calculate the other $(n + 1)$th iterates. Let us say that the matrix of the equations under consideration has the 'Property A', and the equations have been

ordered in a manner so as to give a matrix of the form (5.58) or (5.59). Then a different ordering of the equations is said to be 'consistent with' the form (5.58) or (5.59), when the $(n + 1)$th iterative values for the two orderings are identical for $n = 0, 1, 2, ...$, assuming the initial inputs to be the same. This means that the components of $\mathbf{u}$ corresponding to the diagonal sub-matrix $\mathbf{D}_{k-1}$ must be evaluated before the components of $\mathbf{u}$ corresponding to the sub-matrix $\mathbf{D}_k$. A scanning of the mesh points that is consistent with such a form as (5.58) or (5.59) is defined to be consistent. So the matrices with these forms have the Property A and are consistently ordered. For the 'Property A' matrices which are consistently ordered, the eigenvalues, λ, of the S.O.R. iteration matrix $\{(\mathbf{I} - \alpha\,\mathbf{L})^{-1}\,[\alpha\,\mathbf{U} - (\alpha - 1)\,\mathbf{I}]\}$ are related to the eigenvalues, μ, of the Jacobi iteration matrix $(\mathbf{L} + \mathbf{U})$ by the equation

$$(\lambda + \alpha - 1)^2 = \lambda^2\alpha^2\mu^2, \quad \text{if } |\mu| < 1 \tag{5.60}$$

We shall now give some examples of the consistent ordering. Some of these for the 5-point formula on a square mesh are given by:

1. scanning the successive mesh lines in the same direction;
2. scanning the successive diagonals through the mesh points in the same direction;
3. labelling the mesh points black and white as on a chessboard, and taking all the white equations before the black ones.

Or, in general, we take any two adjacent mesh points P and Q, and put an arrow in the direction P to Q, if P precedes Q in the ordering. Then the ordering is consistent if, on moving once round every square mesh, we move with the arrow on any two sides and against it on the other two sides.

Further details of the consistent ordering will be found in Forsythe & Wasow, Fox, Varga, and Smith.

Finally, though we have mentioned about the errors, we have not discussed them in any depth. For detailed analysis of errors, the readers are recommended to Wilkinson, Fox, Walsh, and Fox and Parker.

5.6 FINITE ELEMENT METHOD

In the finite difference method, the regularity of the discretization mesh is one of the great advantages of programming. But this also at times can produce excessive information. Hence it is a good idea to use larger mesh in the region where no critical information is required, and finer mesh in those parts where detailed information is necessary. The finite difference meshes can thus be graded as required, but the computation becomes complicated for the regions where the meshes of different sizes meet.

The possibility of a completely free topology for the meshes is achieved by the finite element method, which is based on the energy distribution and not on the differential equation describing the equilibrium condition. This method is a variational one and one of the most practical ones for solving the boundary value problems. In this method, the partial differential equation of the field problem is formulated as an 'energy functional' which is then minimized to obtain the solution to the P.D.E.

Hence, in order to solve a boundary value problem, the problem must be formulated in variational terms as a functional; and it must be minimized with respect to a trial solution or a set of trial solutions. The necessary condition for obtaining the functional minimum is that the Euler equation of the functional is the P.D.E. of the field problem.

5.6.1 Functional and Its Extremum

A functional is an operation which assigns a number to each of a certain class of functions. For instance, if we define ϕ to be a set of functions of two variables (x, y) subject to the condition $\phi = f(s)$ on S, where S is the surface bounding the region R and is continually differentiable, then any quantity such as $\Im$ which takes a specific numerical value corresponding to each function in the set, is said to be a 'functional' on the set ϕ. It is a function of a function.

We shall now consider a functional in a single independent space variable, and then subsequently extend it to two space variables.

Let us consider the maximization or the minimization of an integral of the form

$$\Im = \int_a^b F(x, u, u')dx \tag{5.61}$$

subject to the conditions

$$u(a) = A,\ u(b) = B \tag{5.62}$$

where a, b, A, and B are constants. We also assume that F has continuous second-order derivatives with respect to its three arguments, and that the unknown function $u(x)$ possesses two derivatives everywhere in (a, b).

Next, we shall consider a one-parameter family of admissible functions which includes $u(x)$ of the form

$$u(x) + \varepsilon\eta(x) \tag{5.63}$$

where $\eta(x)$ is an arbitrary twice-differentiable function which vanishes at the end-points (a, b). So,

$$\eta(a) = \eta(b) = 0 \tag{5.64}$$

Also ε is a parameter which is a constant in any one function in the set, but varies from one function to another. The increment $\varepsilon\eta(x)$ represents the difference between the varied function and the actual function, and is called a variation of $u(x)$.

If $u(x)$ is replaced by $u(x) + \varepsilon\eta(x)$, then the integral assumes the form

$$\Im(\varepsilon) = \int_a^b F(x, u + \varepsilon\eta, u' + \varepsilon\eta')dx \tag{5.65}$$

It is obvious that the maximum value of the integral occurs when $\varepsilon = 0$, i.e. when the variation of u is zero. Hence

$$\frac{d\Im(\varepsilon)}{d\varepsilon} = 0 \quad \text{when } \varepsilon = 0 \tag{5.66}$$

Now,

$$\frac{d\Im(\varepsilon)}{d\varepsilon} = \int_a^b \left[\frac{\partial F\{x, u + \varepsilon\eta, u' + \varepsilon\eta'\}}{\partial(u + \varepsilon\eta)} \frac{\partial(u + \varepsilon\eta)}{\partial\varepsilon} + \frac{\partial F\{x, u + \varepsilon\eta, u' + \varepsilon\eta'\}}{\partial(u' + \varepsilon\eta')} \frac{\partial(u' + \varepsilon\eta')}{\partial\varepsilon} \right] dx$$

$$= \int_a^b \left[\frac{\partial F\{x, u + \varepsilon\eta, u' + \varepsilon\eta'\}}{\partial(u + \varepsilon\eta)} \eta + \frac{\partial F\{x, u + \varepsilon\eta, u' + \varepsilon\eta'\}}{\partial(u' + \varepsilon\eta')} \eta' \right] dx \tag{5.67}$$

Setting $\varepsilon = 0$, we get

$$\frac{d\Im(\varepsilon)}{d\varepsilon} = \int_a^b \left[\left(\frac{\partial F}{\partial u} \right) \eta(x) + \left(\frac{\partial F}{\partial u'} \right) \eta'(x) \right] dx = I'(0) = 0 \tag{5.68}$$

The second term of the above definite integral in (5.68) can be transformed by integration by parts as below:

$$\int_a^b \left(\frac{\partial F}{\partial u'} \right) \eta'(x) dx = \left[\left(\frac{\partial F}{\partial u'} \right) \eta(x) \Big|_a^b - \int_a^b \frac{\partial}{\partial x} \left(\frac{\partial F}{\partial u'} \right) \eta(x) dx \right] \tag{5.69}$$

The first term of the right-hand side of the above equation is zero, since

$$\eta(a) = \eta(b) = 0$$

Substituting in Eq. (5.68), we get

$$\Im'(0) = - \int_a^b \frac{\partial}{\partial x} \left(\frac{\partial F}{\partial u'} \right) \eta(x)\, dx + \int_a^b \left(\frac{\partial F}{\partial u} \right) \eta(x)\, dx = 0 \tag{5.70}$$

This is known as the Euler–Lagrange equation.

Or

$$\int_a^b \left[\frac{\partial}{\partial x} \left(\frac{\partial F}{\partial u'} \right) - \left(\frac{\partial F}{\partial u} \right) \right] \eta(x)\, dx = 0 \tag{5.71}$$

From this equation, it can be deduced that the integrand must be zero, such that

$$\frac{\partial}{\partial x} \left(\frac{\partial F}{\partial u'} \right) - \frac{\partial F}{\partial u} = 0 \tag{5.72}$$

This expression is called the 'Euler equation of the functional', defined by the integral Eq. (5.61) subject to the boundary conditions specified by Eqs. (5.62).

5.6.2 Functional in Two Variables

Next we consider a functional in two variables, by a surface integral of the form

$$\Im = \iint_R F(x,\ y,\ \phi,\ \phi_x,\ \phi_y) dx\, dy \tag{5.73}$$

where

$$\phi_x = \frac{\partial \phi}{\partial x}, \qquad \phi_y = \frac{\partial \phi}{\partial y} \tag{5.74}$$

We next change the function $\phi(x, y)$ to its new value

$$\phi(x, y) + \varepsilon \eta(x, y) \tag{5.75}$$

The change $\varepsilon \eta(x, y)$ in $\phi(x, y)$ is called the variation of ϕ, and is denoted by

$$\delta\phi = \varepsilon \eta(x, y) \tag{5.76}$$

Corresponding to this change in ϕ, the functional will assume a value

$$\mathfrak{I}(\varepsilon) = \iint_R [F\{x, y, \phi + \varepsilon\eta, \phi_x + \varepsilon\eta_x, \phi_y + \varepsilon\eta_y\}]\, dx\, dy \tag{5.77}$$

Differentiating $\mathfrak{I}(\varepsilon)$ with respect to ε, we get (as before)

$$\begin{aligned}\frac{d\mathfrak{I}(\varepsilon)}{d\varepsilon} &= 0 \\ &= \iint_R \left[\left\{ \frac{\partial F(x, y, \phi + \varepsilon\eta, \phi_x + \varepsilon\eta_x, \phi_y + \varepsilon\eta_y)}{\partial(\phi + \varepsilon\eta)} \right\} \eta + \left\{ \frac{\partial F(x, y, \phi + \varepsilon\eta, \phi_x + \varepsilon\eta_x, \phi_y + \varepsilon\eta_y)}{\partial(\phi_x + \varepsilon\eta_x)} \right\} \eta_x \right. \\ &\quad \left. + \left\{ \frac{\partial F(x, y, \phi + \varepsilon\eta, \phi_x + \varepsilon\eta_x, \phi_y + \varepsilon\eta_y)}{\partial(\phi_y + \varepsilon\eta_y)} \right\} \eta_y \right]\end{aligned} \tag{5.78}$$

Setting $\varepsilon = 0$, we get

$$\frac{d\mathfrak{I}(\varepsilon)}{d\varepsilon} = \mathfrak{I}'(0) = \iint_R \left[\left(\frac{\partial F}{\partial \phi} \right) \eta + \left(\frac{\partial F}{\partial \phi_x} \right) \eta_x + \left(\frac{\partial F}{\partial \phi_y} \right) \eta_y \right] dx\, dy \tag{5.79}$$

The second and the third terms can be expanded by integration by parts, i.e.

$$\iint_R \left\{ \left(\frac{\partial F}{\partial \phi_x} \right) \eta_x \right\} dx\, dy = \oint_S \left(\frac{\partial F}{\partial \phi_x} \right) \eta\, dy - \iint_R \frac{\partial}{\partial x} \left(\frac{\partial F}{\partial \phi_x} \right) \eta\, dx\, dy \tag{5.80}$$

$$\iint_R \left\{ \left(\frac{\partial F}{\partial \phi_y} \right) \eta_y \right\} dx\, dy = \oint_S \left(\frac{\partial F}{\partial \phi_y} \right) \eta\, dx - \iint_R \frac{\partial}{\partial y} \left(\frac{\partial F}{\partial \phi_y} \right) \eta\, dx\, dy \tag{5.81}$$

Substituting from Eqs. (5.80) and (5.81) in Eq. (5.79), we get

$$\begin{aligned}\frac{d\mathfrak{I}(\varepsilon)}{d\varepsilon} &= 0 \\ &= -\iint_R \left\{ \frac{\partial}{\partial x} \left(\frac{\partial F}{\partial \phi_x} \right) + \frac{\partial}{\partial y} \left(\frac{\partial F}{\partial \phi_y} \right) - \left(\frac{\partial F}{\partial \phi} \right) \right\} \eta(x, y)\, dx\, dy \\ &\quad + \oint_S \left\{ \left(\frac{\partial F}{\partial \phi_x} \right) dy + \left(\frac{\partial F}{\partial \phi_y} \right) dx \right\} \eta(x, y)\end{aligned} \tag{5.82a}$$

Removing $\eta(x, y)$ from the above equation, we have

$$\iint_R \left\{ \frac{\partial}{\partial x} \left(\frac{\partial F}{\partial \phi_x} \right) + \frac{\partial}{\partial y} \left(\frac{\partial F}{\partial \phi_y} \right) - \left(\frac{\partial F}{\partial \phi} \right) \right\} dx\, dy - \oint_S \left\{ \left(\frac{\partial F}{\partial \phi_x} \right) dy + \left(\frac{\partial F}{\partial \phi_y} \right) dx \right\} = 0 \tag{5.82b}$$

The first term of Eq. (5.28b) is the Euler-Lagrange equation, and the second term comes out equal to zero. We have thus split the original area integral of $\left. \frac{d\mathfrak{I}(\varepsilon)}{d\varepsilon} \right|_{\varepsilon=0}$ over the domain of integration R into the boundary S, into the two integrals one of which is for the region R excluding the boundary, and the other a boundary integral S.

Equation (5.82b) represents the condition for the extremum (i.e. the maximum or the minimum) of the functional $\Im$. In order to show that the functional is a minimum, we have to establish the relationship

$$\frac{d^2\Im(\varepsilon)}{d\varepsilon^2} > 0 \tag{5.83}$$

Thus the necessary conditions for the functional minima are:

(a) that the first variation or the derivative of the functional is zero (necessary condition);
(b) that the second variation is greater than zero.

In most of the cases, the condition (b) is satisfied when the condition (a) is satisfied. Thus the 'necessary condition' for an extremum is also the sufficient condition for a minimum. So it will be sufficient for us to consider the first variation of the functional.

If now in Eq. (5.82b), the boundary integral is set to zero, the condition for a functional minimum becomes

$$\iint_R \left\{ \frac{\partial}{\partial x}\left(\frac{\partial F}{\partial \phi_x}\right) + \frac{\partial}{\partial y}\left(\frac{\partial F}{\partial \phi_y}\right) - \left(\frac{\partial F}{\partial \phi}\right) \right\} dx\, dy = 0 \tag{5.84}$$

Hence the integrand of Eq. (5.84) must be made equal to zero, which gives the result

$$\left\{ \frac{\partial}{\partial x}\left(\frac{\partial F}{\partial \phi_x}\right) + \frac{\partial}{\partial y}\left(\frac{\partial F}{\partial \phi_y}\right) - \left(\frac{\partial F}{\partial \phi}\right) \right\} = 0 \tag{5.85}$$

This is the Euler equation.

The effects of setting the boundary integral to zero will be examined later. At this stage it can be concluded that the functional is a minimum if the Euler equation is satisfied. In general, this condition is satisfied if the Euler equation of the functional is the same as the partial differential equation of the field problem.

5.6.3 Functional for Electrostatic Fields

It is possible, starting from the Poynting vector (which we shall study later) for the energy distribution in the field, that the functional can be written in terms of the electrostatic potential V and the source charge density ρ_C (if there is any in the specified region). Thus,

$$\Im = \iiint_{\text{vol}} \left\{ V\rho_C - \left(\frac{\varepsilon}{2}\right)|\nabla V|^2 \right\} d(\text{vol}) \tag{5.86}$$

where ε is the permittivity of the medium.

The proof of this equation will follow the chapter on the Poynting vector and the energy transfer. If this functional is substituted in the Euler Eq. (5.85), it reduces to the Poisson's equation for the electrostatics (in which ϕ is replaced by V). Hence the minimization of this functional will give us the solution of the requisite field problem.

5.6.4 Functional and the Boundary Conditions

We next show that the boundary conditions are also satisfied in the process of minimization of

the functional. We consider the contour integral term of the Euler-Lagrange equation and consider its differential which is then set to zero, i.e.

$$\oint_S \left\{ \left(\frac{\partial F}{\partial \phi_x} \right) \cos \phi + \left(\frac{\partial F}{\partial \phi_y} \right) \sin \phi \right\} \delta\phi \, ds = 0 \tag{5.87}$$

where δx and δy have been replaced by ds and the angle which ds makes with the x- and the y-axes. By suitably choosing the value of F, this integral can be expressed as

$$\oint_S \left\{ \left(\frac{\partial F}{\partial \phi_x} \right) \cos \phi + \left(\frac{\partial F}{\partial \phi_y} \right) \sin \phi \right\} \delta\phi \, ds = \oint_S \left(\frac{\partial \phi}{\partial n} \right) \delta\phi \, ds \tag{5.88}$$

Hence, if (a) $\delta\phi = 0$, or (b) $(\partial\phi/\partial n) = 0$, the definite integral vanishes, i.e. (a) $\delta\phi = 0$ condition makes the boundary an equipotential line; and (b) $(\partial\phi/\partial n) = 0$ condition implies that the normal derivative of the potential is zero on this boundary.

Thus in the process of the functional minimization, we have obtained the Euler equation of the functional, and the boundary conditions are satisfied automatically whilst the specified potentials are imposed on the boundaries in the process of solving.

5.6.5 Functional Minimization

To minimize the functional over the region, it is necessary to have discrete representation of the variable (i.e. the potential) over the entire region. Hence the whole region can be divided into a number of elements in any desired manner ensuring that all the material interfaces (if there are any in the specified problem) coincide with the element sides. In practice, the simplest elements are triangular in shape. This process ensures the continuity of the normal boundary conditions across the interfaces. The number, the shape, the size of these elements need not be restricted in any manner. An approximate distribution of the potential is assumed within the element, which (the potential) is a function of the coordinates of the nodes of the element and the nodal values of the potentials at their vertices. Thus we have an approximate distribution of the potential V (or ϕ). These potentials are varied until the functional given by Eq. (5.86) reaches its minimum. The resulting potential distribution gives the best possible solution in the least square sense, since the functional contains terms involving the second order of the potential V. A variety of approximate descriptions for V within the element are possible. In triangular elements, the linear variations of V inside would give (and have given) sufficiently accurate results. For this, the potential V within the triangle lmn (Figure 5.13) is given by

$$V(x, y) = ax + by + c \tag{5.89}$$

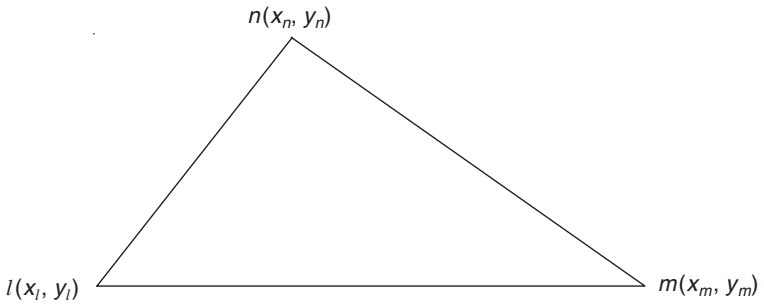


Figure 5.13 A typical element used in F.E.M..

The constants a, b, c can be obtained in terms of the nodal values of V and their coordinates to give

$$V(x, y) = \sum_{i=l,m,n} \left(\frac{p_i + q_i x + r_i y}{2\Delta} \right) V_i \tag{5.90}$$

where Δ = the area of the triangle whose vertices are l, m, n

$p_i = x_m y_n - x_n y_m$
$q_i = y_m - y_n$
$r_i = x_n - x_m$
V_i = the potential at the node i.

The electric flux density within the triangle is given by

$$D_x = -\varepsilon \left(\frac{\partial V}{\partial x} \right) = -\varepsilon \sum_{i=l,m,n} \frac{q_i V_i}{2\Delta}$$

$$D_y = -\varepsilon \left(\frac{\partial V}{\partial y} \right) = -\varepsilon \sum_{i=l,m,n} \frac{r_i V_i}{2\Delta} \tag{5.91}$$

Thus the flux density and also the permittivity inside a triangle is constant, for a first-order element. Equation (5.89) can then be written as

$$V(x, y) = \sum_{i=l,m,n} c_i V_i \tag{5.92}$$

where c_l, c_m, c_n are the area coordinates of the nodes l, m, n given by

$$c_l = \frac{p_i + q_i x_1 + r_i y_1}{2\Delta} \tag{5.93}$$

These area coordinates* are also called the shape functions. Expressing the polynomial in terms of these functions makes it more convenient for further operation. Some of the important properties of these functions are:

1. These functions allow the interpolation of the potential inside the element in terms of the nodal values.
2. At any given vertex, the shape function corresponding to it assumes a value of unity, and zero at other vertices. Thus the magnitudes of these functions alternate between zero and one over the element.
3. The sum of the shape functions for any element is always zero.

Similar equations have to be written for all the triangles, thus defining V and **D** in terms of the nodal values of V at the vertices of the triangles. The functional involves integration of the area and so can be expressed as integrating over each element and then summing all of them to obtain the total integral. So we write:

$$\Im = \sum_{k=1,2,3,\ldots,N} \iint \left[\left(-\frac{\varepsilon}{2} \right) \left(D_{xk}^2 + D_{yk}^2 \right) + \rho_C V_k \right] dS_k \tag{5.94}$$

* Such a three-axis coordinate system (non-orthogonal) for a two-dimensional problem is called 'areal coordinate system', which is a special case of the general homogeneous coordinate system. (For details, refer to Askwith, E.H., *Analytical Geometry of the Conic Sections.*)

where

N = the number of the elements
dS_k = area element of the kth element.

Here each triangle contributes one term to the integral. The flux density **D** within each element is constant and defined in terms of the nodal values of V at the vertices. Hence the minimization of the energy functional is carried out by setting the first derivative of the functional with respect to every nodal value to zero, so that

$$\frac{\partial \mathfrak{I}}{\partial V_k} = \sum_{k=1,2,\ldots,M} \frac{\partial \mathfrak{I}}{\partial V_k} = 0 \tag{5.95}$$

where

k = node number varying from 1 to M
M = the number of the nodes.

The terms of Eq. (5.94) can be expressed in terms of the polynomials for **D** and V, by considering the relevant derivatives of the expression for the functional also expressed in terms of the nodal polynomials. At the end, for each triangular element, we write the three corresponding equations in matrix form which in abbreviated manner can be written down as

$$[s]\ [v] + [t]\ [v] = [\rho] \tag{5.96}$$

Similar equations are formed for every element in the region, and all these matrices are assembled to form a global matrix as

$$[S]\ [V] + [T]\ [V] = [P] \tag{5.97}$$

In Eq. (5.97), the contributions from each element connected to the common node are added to the respective terms in the global matrix [S]. The matrix [T] is known, and is formed from the area of the elements and the material properties. On the right-hand side, the source matrix is assembled by adding the charge distribution at each node. The resultant matrix [S] has the following properties:

(a) The matrix is symmetric and is of the order $m \times m$, where m is the total number of the nodes.
(b) It is sparse and band-structured, i.e. with nonzero elements near the diagonal. This is because all the nodes are not connected with each other.
(c) The matrix is singular, since all the elements of any row add up to zero, thereby resulting the determinant of the matrix to add up to zero.

This means that the equations (5.96) do not give a unique solution. The reason for this is that no boundary conditions have been imposed explicitly so far. All the boundaries are implicitly assumed to be of Neumann type. Thus for such a set of equations, a number of solutions are possible, each differing from the other by a constant. Hence for a unique solution, it is necessary to prescribe the potential at least at one point. Imposition of such a boundary is called the *Dirichlet condition*, which we now discuss.

All the elements of the rows and the columns corresponding to the node with the prescribed potential are made equal to zero, except the diagonal element which is put down as 1.0. The prescribed potential replaces the source term for the node, while the other elements of the column matrix also get modified. The matrix [S], after the imposition of the Dirichlet boundary condition, does not remain singular and a unique solution is obtained. Again it must be remembered that the

Neumann type boundary condition does not have to be imposed explicitly as it is already built in the formulation of the functional. The resultant set of the simultaneous equations represented by the matrix Eq. (5.97) are usually solved by the Gaussian Elimination method.

5.7 GENERAL COMMENTS

We have discussed a number of methods for solving the electrostatic field problems. But it should be remembered that this is not a complete list. We have mentioned only the more important ones and those which have been used more commonly. It should also be noted that the operational equations for the electrostatics satisfy other types of fields, i.e. the Laplace's and the Poisson's equations are also applicable to the magnetic fields, the electric current fields, the steady-state heat flow, etc. So the methods discussed here would be applicable equally to the problems pertaining to these other types of fields. For example, the conducting paper modelling method, which has been briefly mentioned here, has been very widely used for solving the magnetostatic problems. Such methods are going to be discussed in depth in pertinent places in this book. The method of images has also been very widely used in the magnetic field problems in connection with the representation of the boundaries of the problems of electrical machines, particularly in cylindrical geometry. Such aspects will be dealt with in relevant places.

Another point brought to the attention of the readers is about the numerical methods. These methods are also equally applicable to the other types of fields which satisfy the same and similar operating field equations, and hence their capabilities to solve such problems will be highlighted in due course. So far we have explained a few finite difference methods (i.e. Jacobi, Gauss-Seidel, S.O.R., etc.), but this is not a comprehensive list. There are other F.D. methods, which are even more efficient and faster than those described here, for instance, 'the alternating direction implicit method', which have not been included in this section. Readers interested in such methods are recommended books on this subject by Forsythe & Wasow, Varga, Ames, and some of their original papers on the subject. Much work remains to be done in the use of such methods to maximize the utilization of their capabilities.

Next coming to the use of the finite element method, there is a tendency amongst some users to regard this method as a more successful rival of the finite difference methods. This is a very restricted viewpoint; and in fact it should be appreciated that the two methods are complementary. In fact, there has been some significant work done in comparing the efficacy of the two methods, and attempts have been made to identify the areas of preferential use of the either of the two methods. For example, there are areas of electrostatic field problems (and also heat-transfer problems), where the F.D. methods are still found to be preferable, while the magnetic and the electromagnetic field problems are being extensively solved by using the finite element method. This has been the situation with mostly the two-dimensional problems. But with the advent of more and more powerful and faster computers, the three-dimensional problems are also being attempted; while the finite element methods are being used to solve three-dimensional electrostatic and magnetostatic problems, their use in three-dimensional multi-region eddy-current problems remains questionable (at least in the minds of mathematicians). However a detailed discussion of all pertinent points would be the subject matter of specialized textbooks, and cannot be covered in an elementary level discussion such as here.

There are other methods, both approximate as well as analytical, which we have not discussed here. Some of these methods are: Integral equation method, Moment method, Roth's method, and so on. Proper discussion of such methods would be the subject matter of advanced level specialized texts on the relevant topics.

PROBLEMS

5.1 Two point charges, ±1 μC, are located 3 m apart in air. Determine graphically, the location of the equipotential lines in a plane passing through the point charges.

Hint: Initially draw the component equipotential circles for each isolated charge at intervals of 1000 V from 2000 to 10,000 V. The intersection points of the two sets of concentric circles will give the locations of the resulting equipotentials.

5.2 In Section 5.2.2, it has been proved that the scalar variable Φ satisfies the Laplace's equation; starting from Eq. (5.1), show that the orthogonal scalar variable Θ also satisfies the Laplace's equation.

5.3 In cylindrical polar coordinates, starting from the Laplace's equation, show that its finite-difference approximation is:

$$
\begin{aligned}
(\delta r)^2 \nabla^2 V \simeq & \frac{2r+\delta r}{2r} V_{r+\delta r,\phi,z} + \frac{2r-\delta r}{2r} V_{r-\delta r,\phi,z} \\
& + \left(\frac{\delta r}{r\delta\phi}\right)^2 \left(V_{r,\phi+\delta\phi,z} + V_{r,\phi-\delta\phi,z}\right) \\
& + \left(\frac{\delta r}{\delta z}\right)^2 \left(V_{r,\phi,z+\delta z} + V_{r,\phi,z-\delta z}\right) \\
& - 2\left\{1 + \left(\frac{\delta r}{r\delta\phi}\right)^2 + \left(\frac{\delta r}{\delta z}\right)^2\right\} V_{r,\phi,z} \\
= & \ 0
\end{aligned}
$$

5.4 Show that the finite difference approximation for the Laplace's equation in spherical polar coordinates is:

$$
\begin{aligned}
(\delta r)^2 \nabla^2 V \simeq & \frac{r+\delta r}{r} V_{r+\delta r,\theta,\phi} + \frac{r-\delta r}{r} V_{r-\delta r,\theta,\phi} \\
& + \frac{1}{2}\left(\frac{\delta r}{r\delta\theta}\right)^2 (2 + \cot\theta\, d\theta)\left(V_{r,\theta+\delta\theta,\phi} + V_{r,\theta-\delta\theta,\phi}\right) \\
& + \left(\frac{\delta r}{r\sin\theta\,\delta\phi}\right)^2 \left(V_{r,\theta,\phi+\delta\phi} + V_{r,\theta,\phi-\delta\phi}\right) \\
& - 2\left\{1 + \left(\frac{\delta r}{r\delta\theta}\right)^2 + \left(\frac{\delta r}{r\sin\theta\,\delta\phi}\right)^2\right\} V_{r,\theta,\phi} \\
= & \ 0
\end{aligned}
$$

6 Steady Electric Current and Electric Field

6.1 INTRODUCTION

So far we have considered the static electric field, i.e. the field due to the static electric charges. Now we shall consider the situations where the continuous distribution of charges are macroscopically in a steady state of motion. Such a state happens when large aggregates of elemental electric charges are moving at constant speed. When the charges are no longer static, the theory of the field of the static charges, i.e. the Coulomb's law of force, has to be modified. The force between the pairs of moving charges cannot be described completely by the Coulomb's law, and new terms in the equation have to be incorporated as we shall see later. Since, however, our interest is in 'the continuous distribution of charges in steady state of motion', the forces of the new type on the individual elemental charges may be considered negligible. The steady motion of the charges is canalized in suitable conductors. By conductors we mean any substance which contains in it a relatively large number of charges (i.e. electrons) free to move about. Metals which are the most important of the conductors contain in general a large number of valence or 'conduction electrons' which are very loosely connected with the rest of the atoms. In normal conditions, these electrons move at random from one atom to another. But even if a very weak electric field is imposed on the conducting medium, an organized motion of the electrons gets superimposed on the random motion. Such an organized motion of the charged elementary particles is referred to as the 'electric current'. The imposed electric field accelerates the electrons in the direction of the field but sooner or later they collide with other atoms as a result of which the average velocity of the electrons in the directed motion becomes quite small, and due to the collisions the electrons transfer a part of this kinetic energy to the atoms. This transferred energy appears as vibrations of the atoms which represent the thermal energy. The electric current in any conductor is thus associated with the generation of heat which is known as 'Joule's heat'.

So far we have described qualitatively the mechanism of the electric conduction in solid conductors, which is quite different from that in liquids. In a liquid conductor, certain percentage of molecules dissociate into two oppositely charged parts, which are called *ions.* When an electric field is imposed on such a liquid, the positive ions move in the direction of the field and the negative ions move in the opposite direction. The electric current in the liquid conductors is thus made up of two streams of charged particles moving in opposite directions, and generally with different average velocities. As in solid conductors, there would be a generation of heat here as

well. In addition, the electric current in the liquid conductors is associated with chemical changes, a process which we know as *electrolysis.*

The number of free electric charges per unit volume of solid and liquid conductors is extremely large, because there are approximately some 10^{22} to 10^{24} atoms per cubic centimetre of a rigid body. In 10^{-6} cubic centimetres of a metallic conductor, there are approximately 10^{16} to 10^{18} electrons. So macroscopically we can visualize the electric current as the flow of some kind of 'electric fluid'.

6.2 ELECTRIC CURRENT AND CURRENT DENSITY

So far we have tried to visualize and describe the electric current in a qualitative manner. Quantitatively, the electric current, which is the flow of a continuous distribution of electric charges, can be defined by two physical quantities, i.e.

(a) the current density vector **J** which defines the flow of the charges at a point, and
(b) the current or the current intensity I which describes the rate of flow of the electric charges through a macroscopic surface.

Let us now consider a point inside a conductor where there are, say, N free charges per unit volume, each carrying a charge Q, moving with an average velocity **v**. Then at the point under consideration, the current density vector **J** is given by

$$\mathbf{J} = NQ\mathbf{v} \tag{6.1}$$

According to this definition, the current density vector of the negative charges $-Q$ with the average velocities $-\mathbf{v}$ is the same as the current density vector of the positive charges $+Q$ with the average velocities $+\mathbf{v}$, i.e.

$$\mathbf{J}_{-Q} = N(-Q)(-\mathbf{v}) = NQ\mathbf{v} = \mathbf{J}_{+Q} \tag{6.2}$$

Hence a stream of negative charges in one direction is equivalent to the stream of the positive charges $+Q$ moving in the opposite direction. In general, the electric current may be made up of a continuous distribution of different charges Q_1, Q_2, Q_3, ..., Q_n of different numbers N_1, N_2, N_3, ..., N_n per unit volume and with different velocities $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$, ..., $\mathbf{v}_n$, respectively, at the point under consideration. Thus the current density at that point is the vector sum of the partial current densities, i.e.

$$\mathbf{J}_{\text{total}} = N_1Q_1\mathbf{v}_1 + N_2Q_2\mathbf{v}_2 + N_3Q_3\mathbf{v}_3 + \cdots + N_nQ_n\mathbf{v}_n \tag{6.3}$$

Let us next consider a small planar surface δS in a stream of the charged particles which constitute the electric current (Figure 6.1). Let us assume that the current density vector at that point is

$$\mathbf{J} = NQ\mathbf{v}$$

During a small time interval δt, the surface δS will be crossed by the charge δQ which occupied the volume $\delta S\,(\mathbf{v}\delta t)\cos\alpha$ as shown in Figure 6.1. Hence the charge that crosses the surface δS in this time interval δt is

$$\delta Q = NQ\mathrm{v}\delta S\delta t\cos\alpha \tag{6.4}$$

The electric current or the current intensity that passes through the surface is thus defined as the

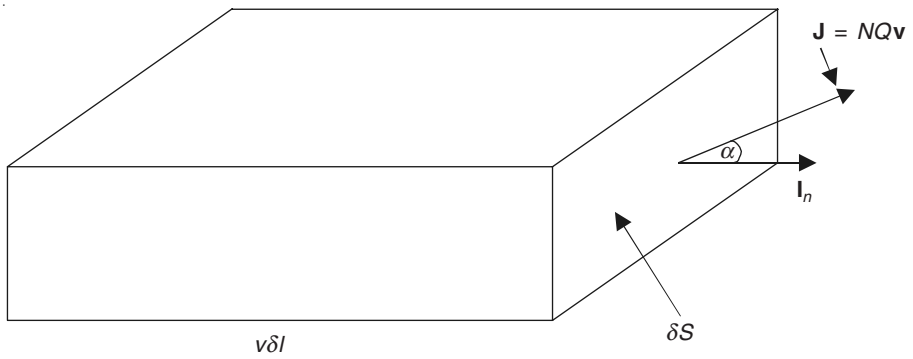


Figure 6.1 A planar surface δS with current across it.

charge crossing the surface per unit time. The symbol for the current is I for the steady time-independent current, and $i(t)$ or i for the time-varying currents. So

$$\delta I = \frac{\delta Q}{\delta t} = NQ\mathrm{v}\delta S \cos \alpha \tag{6.5}$$

Expressing the surface area as a vector,

$$\delta I = \frac{\delta Q}{\delta t} = NQ\mathbf{v} \cdot \delta\mathbf{S} = \mathbf{J} \cdot \delta\mathbf{S} \tag{6.6}$$

So the total current intensity I through an arbitrary surface S is

$$I = \frac{dQ}{dt} = \iint \mathbf{J} \cdot \delta\mathbf{S} \tag{6.7}$$

So the current through a surface may be either positive or negative. Its sign depends on the sign of the positive normal to the surface. By definition, the current I is a scalar quantity, and when we refer to the direction of the current in the wire conductors, it is assumed to be in the direction of the current density vector. By convention, the direction of a current in a metal wire is opposite to the direction of the flow of the free electrons which make up the current.

The unit of current (or the current intensity) is coulomb per second or ampere. Its notation is A which stands for C/s. The lines of the current density vector **J** or the current flow lines are defined as the lines having the property that the **J** vector is tangential to such a line at all the points. So a tubular surface formed by a bunch of current flow lines, would be such that the flux of **J,** i.e. I, is the same through any cross-section of the tube.

6.3 ELECTRIC CURRENT AND ELECTRIC FORCE

We have seen in previous chapters that for a conductor containing free electrons in electrostatic conditions, the electric field **E** is zero and hence the potential must be constant throughout the material of the conductor. But when the current flows, there would exist an electric force **E** within the conducting material and we find that the average local direction of the electron flow is the same as the direction of $-\mathbf{E}$. In many materials, the current traversing the unit area normal to the lines of force is decided by the local value of **E**, and for a large class of conductors, the current is proportional to **E** over a wide range. This is the 'Ohm's law', and the conductors which obey this law are called ohmic or linear. Most metals and aqueous solutions of salts fall in this

category. Books on the electric circuit theory deal with this law and other circuit laws, i.e. Kirchhoff's laws in great detail.

The Ohm's law is usually expressed by the equation

$$V = RI$$

where it is implied that the current flowing in a conductor bears a constant ratio to the potential difference across its ends. The constant R is called the resistance of the conductor and is measured in 'ohms' when V and I are measured in volts and amperes, respectively. It has been found that R is directly proportional to the length and inversely proportional to the cross-sectional area of the conductor. The resistance of a unit cube of the material is its characteristic at any given temperature and is known as its 'resistivity' (= ρ). So we can express the Ohm's law as follows:

Potential difference between the ends of the conductor element,

$$V = El$$

where l is the length of the conductor.

Resistance between the ends of the conductor,

$$R = \frac{\rho l}{S}$$

where S is the cross-sectional area of the conductor.

$\therefore$ The Ohm's law equation $V = RI$ becomes

$$El = \left(\frac{\rho l}{S}\right) I \qquad \text{or} \qquad E = \rho\left(\frac{I}{S}\right) = \rho J \qquad \text{[from Eq. (6.7)]} \tag{6.8}$$

So we find that the current density vector **J** has the same direction as the electric field intensity **E**.

6.4 THE CONSERVATION OF CHARGE (THE EQUATION OF CONTINUITY)

Since the charges are indestructible, if there is an inflow of charges in some part of the surface of a conductor, then there would be an equivalent outflow elsewhere or there would be an accumulation of charge upon the conductor. Under the steady-state conditions, ruling out the possibility of charge accumulation, the total inflow of the current to any conductor must equal the total outflow, a condition which must be true for any part of the conductor. This is the 'Principle of conservation of charge'.

Let us consider a closed surface S, enclosing at any instant of time, a total charge Q. The charge Q is the part of a cloud of charges of density ρ_C which is a function of position and time. This charge cloud is moving in an arbitrary manner, and hence the total current $i_S(t)$ through S with respect to its outward normal is

$$i_S(t) = \oiint_S \mathbf{J} \cdot d\mathbf{S} \tag{6.9}$$

where **J** is the current density vector at $d\mathbf{S}$ (Figure 6.2). This current $i_S(t)$ represents the positive charge which crosses S per unit time in the positive (outward) direction. But according to the

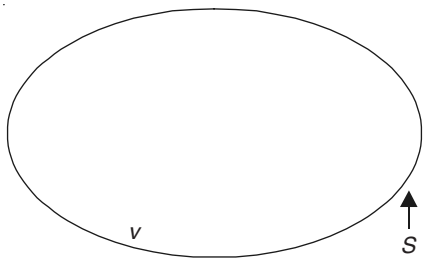


Figure 6.2 A charge cloud enclosed in a closed surface *S*, enclosing volume *v*.

principle of conservation of charges, the outflow of charges must equal the decrease of charges enclosed in *S*, i.e.

$$i_S(t) = \frac{dQ_{\text{crossing}}}{dt} = -\frac{dQ_{\text{inside}}}{dt} = -\frac{d}{dt}\iiint_v \rho_C \, dv \tag{6.10}$$

Combining Eqs. (6.9) and (6.10), we get

$$\oiint_S \mathbf{J} \cdot d\mathbf{S} = -\frac{d}{dt}\iiint_v \rho_C \, dv \tag{6.11}$$

By using the Gauss' divergence theorem, the surface integral can be converted into a volume integral, i.e.

$$\iiint_v \text{div } \mathbf{J} \, dv = -\iiint_v \left(\frac{d\rho_C}{dt}\right) dv$$

or

$$\iiint_v \left\{\nabla \cdot \mathbf{J} + \left(\frac{d\rho_C}{dt}\right)\right\} dv = 0 \tag{6.12}$$

where *v* is an arbitrary volume, and so in the differential form

$$\nabla \cdot \mathbf{J} + \frac{d\rho_C}{dt} = 0 \tag{6.13}$$

This is the differential form of the equation of continuity. If the electric current does not change with time, and the charge density is also independent of time, then $(d\rho_C/dt) = 0$, and hence

$$\nabla \cdot \mathbf{J} = 0 \tag{6.14}$$

Under this condition, the current density vector has no source, and charges cannot be created at any point in a steady current flow. Equation (6.13) is the mathematical form of the principle of conservation of charges, and states that the charges are neither created nor destroyed in macroscopic quantities.

6.5 ANALOGY BETWEEN ELECTRIC CURRENT AND ELECTRIC FLUX

Let us consider a small area δS inside a current-carrying conductor (Figure 6.3) in which the current flow δI is given by

$$\delta I = J(\cos\,\theta)\,\delta S \tag{6.15}$$

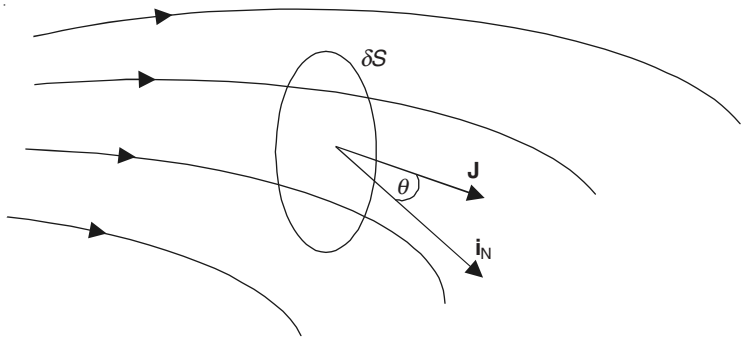


Figure 6.3 Current flow in conductor.

where θ is the angle between the normal to δS and the local direction of the current density vector **J.** But the right-hand side of this equation is the flux of **J** across the surface element δS. And by the principle of conservation of charges discussed in Section 6.4, this quantity is zero.

We should also remember that the electric field vectors **E** and **D** have the same property, i.e. $\nabla \cdot \mathbf{E}$ is zero and $\nabla \cdot \mathbf{D}$ is zero in charge-free regions (and that the regions have uniform permittivity). We have used this analogy in a number of experimental methods of solving field problems, like the conducting paper method and the electrolytic tank method (Chapter 5, Sections 5.3.2 and 5.3.1) to solve the electrostatic field problems which satisfy the Laplace's equation. So now we can establish the scaling relationships between the current flow analogue and the original electrostatic problem. The analogy between the current and the electric flux is

$$\Psi = \iint_S \mathbf{D} \cdot d\mathbf{S} \qquad \text{and} \qquad I = \iint_S \mathbf{J} \cdot d\mathbf{S} \tag{6.16}$$

and for a closed surface

$$\oiint_S \mathbf{D} \cdot d\mathbf{S} = 0 \qquad \text{and} \qquad \oiint_S \mathbf{J} \cdot d\mathbf{S} = 0$$

Thus **J** and **D** obey the same physical laws.

So we can set up an analogue of a system of perfect conductors in a dielectric medium of permittivity ε, by means of a system of similarly shaped perfect conductors in a conducting medium of resistivity ρ (Figure 6.4). Let the linear dimensions be in the ratio of $\lambda : 1$.

In the original problem, $V = \int_C \mathbf{E} \cdot d\mathbf{l}$, and in the analogue, $V' = \int_C \mathbf{E}' \cdot d\mathbf{l}'$.

$$\therefore \qquad \frac{E'}{V'} = \frac{\lambda E}{V} \tag{6.17}$$

So,

$$D = \varepsilon E = \varepsilon\left(\frac{VE'}{\lambda V'}\right) = \left(\frac{\varepsilon V}{\lambda V'}\right)\left(\frac{J'}{\sigma}\right)$$

$$= \left(\frac{\varepsilon V}{\lambda V'}\right)(\rho J') \tag{6.18}$$

where ρ = the resistivity of the medium of the analogue = $1/\sigma$, σ being the conductivity which is the reciprocal of the resistivity.

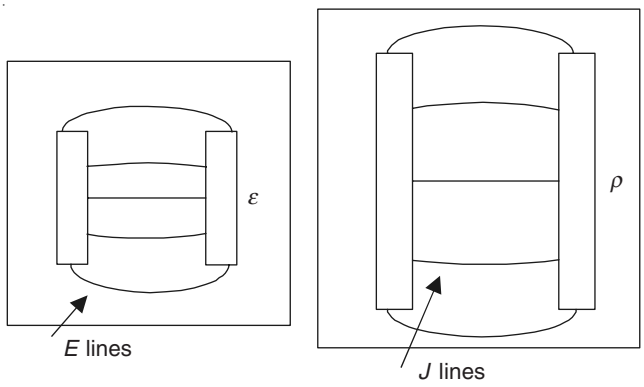


Figure 6.4 Current-flow analogue (either conducting paper or electrolytic tank).

$\therefore$ Flux of **D** (over any area) = λ^2 Flux of $\varepsilon\rho\left(\frac{V}{\lambda V'}\right)J'$, over the corresponding area of the model.

$\therefore$ Total flux of **D** = charge on the electrode Q (say)

and

total flux of $\mathbf{J}'$ = current from the electrode I'

where

$$Q = \lambda\varepsilon\rho\left(\frac{V}{V'}\right)I' \tag{6.19}$$

$\therefore$ Capacitance between the electrodes of the original system,

$$C = \frac{Q}{V} = \lambda\varepsilon\rho\left(\frac{I'}{V'}\right)$$

or

$$C = \lambda\varepsilon\rho G' \tag{6.20}$$

where G' is the conductance between the electrodes of the analogue.

$\therefore$ $$\frac{C}{G'} = \lambda\varepsilon\rho = \text{constant of the analogue} \tag{6.21}$$

6.6 ELECTROMOTIVE FORCE

In the previous section, we have observed a correspondence between the behaviour of the electric current I and the electric flux Ψ. However, it must be understood that this correspondence is conceptually based on mathematical formalism rather than physical. In fact, physically the electric current and the electric flux are quite dissimilar. The current is a flow of the electric charges, and the flux is not a flow of anything—it is in fact a mathematical abstraction. Further to maintain a steady electric current, a continuous source of energy is needed, whereas there is no such requirement for the maintenance of a steady flux. As we have explained before and also based on our everyday experience, we find that when a conducting wire carries an electric current, the wire gets heated because some of the electrical energy supplied to it is being converted to thermal

energy. This behaviour is implicit in the Ohm's law. It can be easily deduced that when a current I flows through a resistance R, the rate of energy loss is I^2R watts. The agency which supplies the energy for maintaining the potential difference across the resistor for the current through it, also supplies this energy. The source can be either a chemical battery or an electric generator. This property of the generator or the battery is called its 'electromotive force', which is commonly abbreviated to 'emf'. The direction of the emf is assumed to be the same as the positive direction of the current flow, and its magnitude is measured by the energy supplied per unit charge delivered. But the available energy per coulomb in the emerging charges at the positive terminal of the battery or the generator is equal to the potential difference through which the charges will fall while travelling round the circuit. Hence the emf defined in terms of energy 'per unit charge' is measured in volts. This energy will be slightly in excess of the P.D. across the terminals of the conductor, as some energy is used up in the source itself. But if the current-carrying circuit is removed, and the P.D. is measured by some instrument which takes negligible current then what we measure would approximate the emf.

What we have discussed qualitatively so far, can thus be expressed as follows. In the source, at any point the charges are acted upon by (a) the electric field of the distributed charges (= $\mathbf{E}$) and (b) the non-electric impressed force $\mathbf{F}_i$ which can be represented as $Q\mathbf{E}_i$, where $\mathbf{E}_i$ is not a real electric force due to the stationary charges. By definition, it gives the correct force on the charges, only when multiplied by the charge.

$\therefore$ The total force $\mathbf{F}_{total}$ on the charges passing through the point under consideration is

$$\mathbf{F}_{total} = Q\mathbf{E} + \mathbf{F}_i = Q(\mathbf{E} + \mathbf{E}_i) \tag{6.22}$$

6.7 POTENTIAL IN THE ELECTRIC CIRCUIT

In an electrostatic field of a group of stationary charges, the work done in carrying a charged particle round a closed path is zero, i.e.

$$\oint_C \mathbf{E} \cdot d\mathbf{l} = 0 \tag{6.23}$$

But this is not true in a current-carrying circuit. For example, let us consider a generator supplying a current to a resistor. In such a circuit, $\mathbf{E}$ is always in the same direction, even in the generator, as otherwise the charges would not keep on moving. So $\oint_C \mathbf{E} \cdot d\mathbf{l}$ cannot be zero, because the sum of the positive contributions itself must be positive. So Eq. (6.23) gets replaced by

$$\mathcal{E} = \int_N^P E_i \, dl_i = -\int_N^P \mathbf{E} \cdot d\mathbf{l} \tag{6.24}$$

that is, the electromotive force $\mathcal{E}$ of a generator is defined as the line integral of the impressed field $\mathbf{E}_i$ from the negative electrode (N) to the positive electrode (P). The impressed field E_i inside the source is exactly equal in magnitude and opposite to the electric field intensity $\mathbf{E}$ due to the distributed charges (Figure 6.5).

The current flow macroscopically coincides with the moving charges in the conducting media and $\mathbf{J}$ is in the direction of the total force $\mathbf{F}_{total}$.

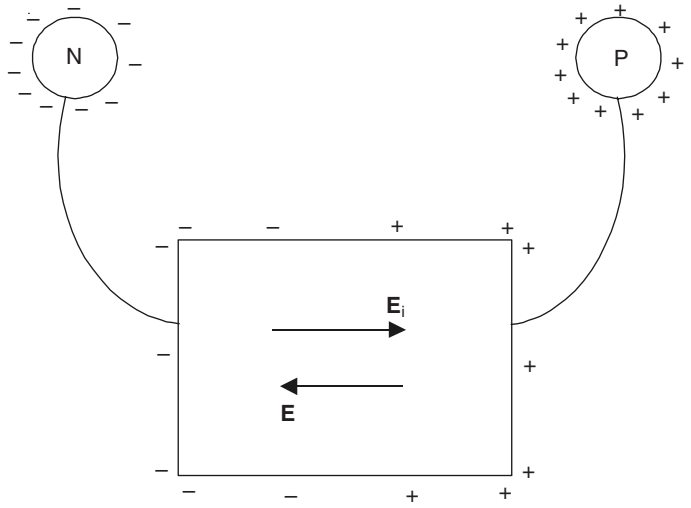


Figure 6.5 An open-circuited source of emf.

$$\therefore \qquad \mathbf{J} = \sigma(\mathbf{E} + \mathbf{E}_i)$$

and outside the generator

$$\mathbf{J} = \sigma\mathbf{E} \tag{6.25}$$

Note: $\oint_C \mathbf{E} \cdot d\mathbf{l}$ in a static field = 0, and

$\oint_C \mathbf{E} \cdot d\mathbf{l}$ in the region of steady current flow = $\mathcal{E}$ (source emf).

6.8 OHM'S LAW AND JOULE'S LAW

Macroscopically, the current flow lines of a steady current flow coincide with the lines of the total force on the moving charges. This is true for all the liquid and most of the solid conductors. In terms of the current density vector **J**, and the total force on the charge, this can be expressed as

$$\mathbf{J} = \sigma(\mathbf{E} + \mathbf{E}_i)$$

where the impressed field is present and

$$\mathbf{J} = \sigma\mathbf{E}, \qquad \text{in absence of the impressed field.}$$

This is Ohm's law and $\sigma (= 1/\rho)$ is the conductivity of the conducting medium and where ρ is the resistivity of the medium as defined before in Section 6.3. The resistivity and the conductivity are the functions of the temperature and can be expressed approximately as

$$\rho(t) = \rho_0\{(1 + \alpha(t - t_0)\} \tag{6.26}$$

where ρ_0 is the resistivity at the reference temperature t_0 and α is referred to as the temperature coefficient.

The unit of the resistivity can be written from Eq. (6.8), i.e.

$$R = \frac{\rho l}{S} \qquad \text{or} \qquad \rho = \frac{RS}{l}$$

So the unit of ρ is ohm-metre^2/metre = ohm-metre = Ω-m (= V-m/A), and the unit of σ is (A/V-m) = mho/m = Siemens/metre = S/m.

Table 6.1 Resistivity and conductivity of some materials

Material	ρ (Ω-m) at 20° C	σ (S/m)
Silver	0.163×10^{-7}	6.15×10^{7}
Copper	0.174×10^{-7}	5.75×10^{7}
Aluminium	0.283×10^{-7}	3.54×10^{7}
Iron	1.00×10^{-7}	1.00×10^{7}
Lead	2.10×10^{-7}	0.48×10^{7}
Sea water	approx. 0.25	approx. 4
Earth	10^{2}–10^{4}	10^{-4}–10^{-2}

Next we consider a conductor having N free charges, each of Q coulombs, per unit volume. And let the electric field inside the conductor be **E** and the average velocity of the free charges be **v**.

∴ The work done by the electric forces in the time-interval δt in moving all the charges inside a small volume element δv is given by

$$\begin{aligned} \delta W_e &= Q\mathbf{E} \cdot (\mathbf{v}\,\delta t) N \delta v \\ &= NQ\mathbf{v} \cdot \mathbf{E}\,\delta v \delta t \\ &= \mathbf{J} \cdot \mathbf{E}\,\delta v \delta t \qquad \text{[from Eq. (6.6)]} \end{aligned} \tag{6.27}$$

As we have explained before, this part of the energy is taken up by the atoms of the conductors due to collisions with free electrons and appears as heat, and so cannot be returned to the electric field. Hence the rate at which the energy of the electric field is transformed into thermal energy (or the power) is

$$\frac{dP}{dv} = \frac{\left(\dfrac{dW_e}{dt}\right)}{dv} = \mathbf{J} \cdot \mathbf{E} \tag{6.28}$$

and so the power absorbed by a volume v of the conductor which is carrying a current is

$$P = \iiint_v \mathbf{J} \cdot \mathbf{E}\, dv \tag{6.29}$$

This is the integral form of the Joule's law. This heating is also known as Joule heating due to the electric current.

6.9 BOUNDARY CONDITIONS

Let us consider a boundary surface between the two conductors of different conductivities σ_1 and σ_2 (Figure 6.6). Applying the continuity equation in the integral form [i.e. Eq. (6.14)] to a coin-shaped Gaussian surface on the interface (for the steady-state condition), we get

$$J_{1n} = J_{2n} \qquad \text{or} \qquad \sigma_1 E_{1n} = \sigma_2 E_{2n} \tag{6.30}$$

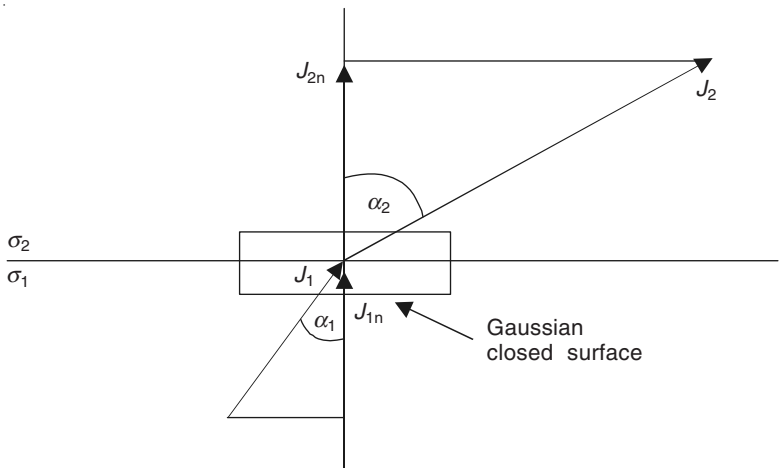


Figure 6.6 Conductor interface.

that is, the normal components of the electric current vector are equal on both the sides of the conductor interface.

When there is no current source on the interface, we can consider a rectangular closed contour and apply the law

$$\mathbf{E} \cdot d\mathbf{l} = 0$$

which gives

$$E_{1t} = E_{2t} \qquad \text{or} \qquad \frac{J_{1t}}{\sigma_1} = \frac{J_{2t}}{\sigma_2} \tag{6.31}$$

that is, the tangential component of the electric field intensity is continuous across the interface. Combining Eqs. (6.30) and (6.31), we get

$$\frac{J_{1t}/J_{1n}}{J_{2t}/J_{2n}} = \frac{\tan \alpha_1}{\tan \alpha_2} = \frac{\sigma_1}{\sigma_2} \tag{6.32}$$

The surface densities of the free charges (= ρ_S) can be evaluated, if the permittivities of the media are known, i.e.

$$\rho_S = D_{2n} - D_{1n} = \varepsilon_2 E_{2n} - \varepsilon_1 E_{1n} \tag{6.33}$$

where ε_2 and ε_1 are the permittivities of the two media, respectively.

6.10 CIRCUIT LAWS

A generator of emf is, in general, a region inside a conductor where the external non-electrical forces act on the free electrical charges. The generators encountered in practical applications have two equipotential terminals called the electrodes. Without going into the mechanism of propelling the electric charges inside such a device, we consider only its emf and its internal resistance.

Also in many practical applications, the current flow is restricted to small pieces of resistive substances. Such pieces are called two-terminal resistors. A number of such elements connected in series or in parallel or in a combination of both, form circuits through which the currents are made to pass from one or more generators. These circuits form the basis of the electric circuit theory. There are different textbooks dealing with the analysis and the algebra of the circuit theory, and

in a textbook on electromagnetism we do not intend to enter into these details. We shall, however, very briefly mention the two basic laws of the circuit theory before we close this chapter.

We consider first the two-terminal generator, when it is not connected to any external circuit, i.e. when it is open-circuited. The impressed field $\mathbf{E}_i$ would cause the charges of the opposite natures to be accumulated around the terminals as shown in the Figure 6.7.

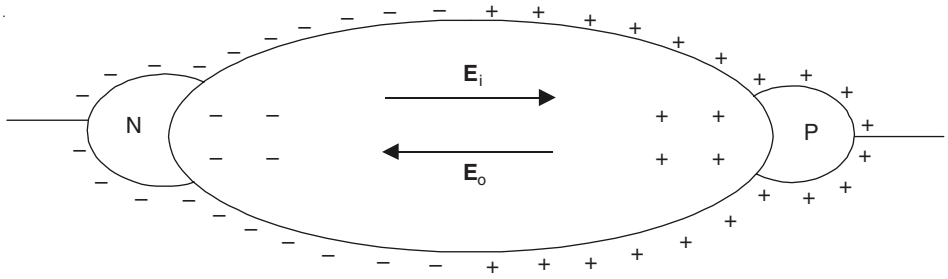


Figure 6.7 Open-circuited two-terminal generator.

Under the equilibrium condition, the electric field intensity $\mathbf{E}_o$ due to these charges would be exactly equal and opposite to $\mathbf{E}_i$. Note that if $\nabla \cdot \mathbf{E}_i \neq 0$, then $\nabla \cdot \mathbf{E}_o \neq 0$ as well. The charges will be accumulated not only on the surface but also inside the generator. The emf equation under open-circuited condition would be

$$\mathcal{E} = V_P - V_N$$

When the generator is connected to an external circuit which draws a current I, the current density being $\mathbf{J}$, the field of the accumulated charges will no longer compensate for $\mathbf{E}_i$, because now at all points $(\mathbf{E}_o + \mathbf{E}_i) = \mathbf{J}/\sigma$, σ being the conductivity inside the generator. The P.D. between the electrodes will be smaller than $\mathcal{E}$. The difference between $\mathcal{E}$ and this present value is $\Delta(V_P - V_N)$ which is necessary to maintain the current I. This P.D. is proportional to I and can be written as

$$\Delta(V_P - V_N) = IR_g$$

where R_g is referred to as the 'internal resistance of the generator'. Thus the P.D. between the electrodes of the generator, when a current I flows through it, is

$$V_P - V_N = \mathcal{E} - IR_g \tag{6.34}$$

On the other hand when a current is forced through the generator in the direction opposite to $\mathbf{E}_i$, the P.D. between the electrodes is larger than the emf of the generator, and we have

$$V_P - V_N = \mathcal{E} + IR_g \tag{6.35}$$

Symbolically, R_g is denoted as shown in Figure 6.8.

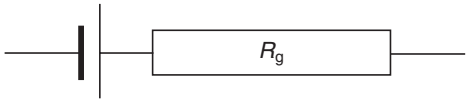


Figure 6.8 Generator with internal resistance R_g.

Now we shall state the two basic laws of the circuit theory.

Kirchhoff's first law. This is really the equation of continuity, specially applied to the nodes of an electrical network. When the continuity equation in integral form, i.e.

$$\oiint_S \mathbf{J} \cdot d\mathbf{S} = 0$$

is applied to a surface S enclosing the node of an electrical network, then the sum of all the currents leaving the node through the conductors meeting at the node must be zero. Thus,

$$\Sigma I = 0 \tag{6.36}$$

For example, in a circuit as shown in Figure 6.9, at the node 1 of the network,

$$-I_1 - I_2 + I_3 + I_4 = 0$$

If there are N nodes in a network, then $(N - 1)$ independent equations for the currents in the branches can be written. The equation for the Nth node is not independent.

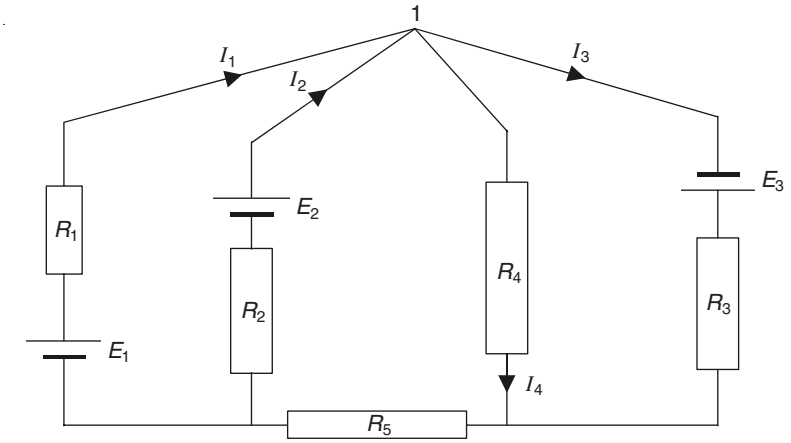


Figure 6.9 An electrical network.

Kirchhoff's second law. Considering a closed circuit as shown in the Figure 6.10, consisting of a single generator and a resistor, we have

$$V_P - V_N = IR$$

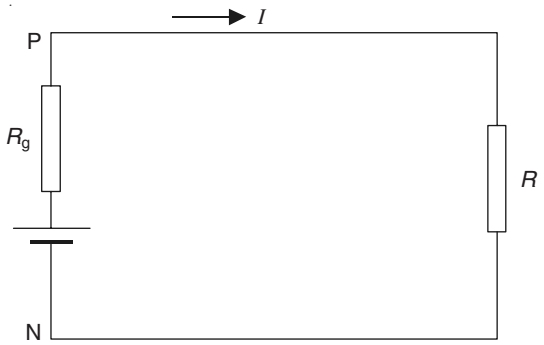


Figure 6.10 A steady current circuit.

and also

$$V_P - V_N = \mathcal{E} - IR_g$$

$$\therefore \qquad \mathcal{E} - IR_g = IR \tag{6.37}$$

This is the Ohm's law for a simple circuit. Generalizing this for any closed contour in any network, this becomes

$$\Sigma\mathcal{E} - \Sigma(IR) = 0 \tag{6.38}$$

All the emfs and all the currents are assumed to be in the adopted positive direction of the contour. This is the Kirchhoff's second law.

6.11 SERIES AND PARALLEL CONNECTION OF RESISTORS

When n resistors are connected end-to-end, they are said to be connected in series. When this circuit is connected to an emf source, the same current flows through all the resistors. Hence the P.D. between the ends of the connection is

$$V = R_1I + R_2I + R_3I + \cdots + R_nI$$

Hence the equivalent resistance R_S is

$$R_S = \frac{V}{I} = R_1 + R_2 + R_3 + \cdots + R_n \tag{6.39}$$

When each of the n resistors is connected to the same two terminals, they are said to be connected in parallel, and by the Kirchhoff's second law

$$R_1I_1 = R_2I_2 = R_3I_3 = \cdots = R_nI_n = V$$

and by Kirchhoff's first law

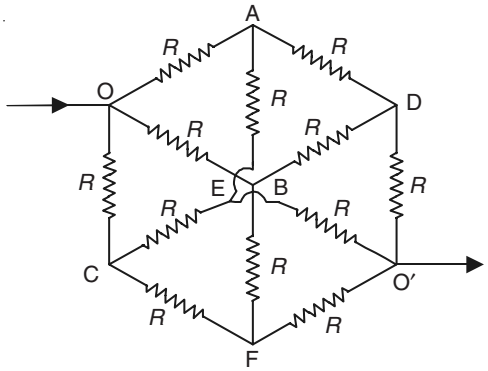
$$I_1 + I_2 + I_3 + \cdots + I_n = I$$

Hence the equivalent resistance is given by

$$\frac{1}{R_P} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots + \frac{1}{R_n} \tag{6.40}$$

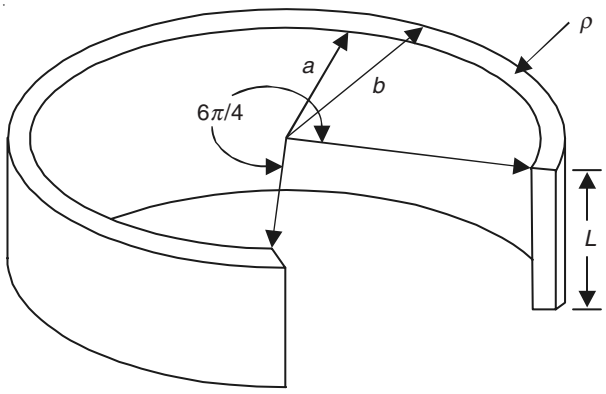
PROBLEMS

6.1 Find the effective resistance of the following circuit.



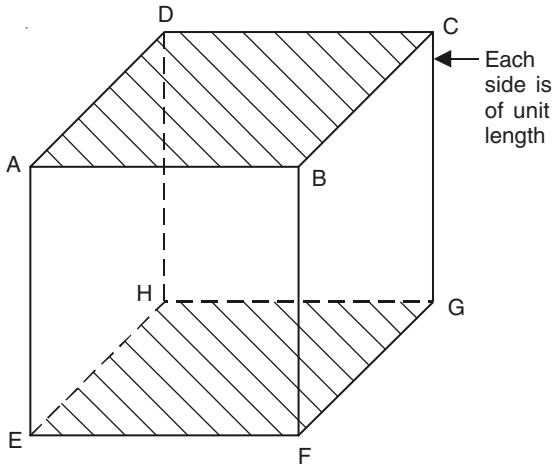
Ans.: $R_{eH} = 5R/6$

6.2 A uniform conducting toroid of rectangular cross-section is cut as shown in the adjoining figure. Assuming that the strip conductors to the toroid are perfect conductors, calculate the resistance offered to the flow of a steady current.



Ans.: $R = 3\rho\pi/\{2L \cdot \ln (b/a)\}$

6.3 Two perfectly conducting metal plates are in contact with a block of conducting material, as shown in the adjoining figure, in which the conductivity of material varies uniformly from σ_1 (on the face ABCD) at one plate to σ_2 at the other (the face EFGH) calculate the resistance between the plates ($\sigma_2 > \sigma_1$).



Ans.: $R = \ln (\sigma_2/\sigma_1)/(\sigma_2 - \sigma_1)$

6.4 What will be the resistance between the plates in the structure of Problem 6.3, when the contacts are on the faces ADHE and BCGF?

Ans.: $R = 2/(\sigma_1 + \sigma_2)$

6.5 A system of 30 conductors, each of equal resistance, is built in the same way as the edges of a dodecahedron. Prove that the resistance of the network between a pair of opposite corners is 7/6 of the resistance of a single conductor.

7 Magnetic Field of Steady Currents in Free Space

7.1 INTRODUCTION

We have seen that the electric force between the two static charges is given by the Coulomb's inverse square law. If however the two charges are in motion, the experimental conditions for the validity of the Coulomb's law do not hold. A correction term is needed for the Coulomb's law equation. The additional force term exerted by a steady current on either a single moving charge or another steady current system is referred to as a 'magnetic force'.

Before we establish the mathematical formalism for the derivation of the magnetic force in a logical manner, i.e. from the static charges which we have studied so far to the systems of moving charges, we shall take a short diversion and have a look at the historical background of the magnetic forces and the magnetism. Historically, however, the science of magnetism began with the study of the permanent magnets. Man first discovered the natural magnets such as lodestone, then observed their north-pointing property, and this was followed by the invention of the compass. Then the possibility of magnetizing the iron bars was observed as was the observation of two kinds of magnetism, i.e. the ends of the magnet bars either attract or repel. The idea, that the behaviour of the compass needle was due to the pole-star, was discredited by William Gilbert, the physician to the Queen Elizabeth-1 of Great Britain. Gilbert showed that such phenomena could be explained by assuming the earth to be a magnet, equivalent to a magnetized sphere.

The first attempts to quantify the science of magnetism were made by Coulomb who also derived the inverse square law for the electrically charged particles. He established that the extremities of the magnets or the 'poles' of the magnets obey the same inverse square law of forces. This was the start of the parallel development of the sciences of the electrostatics and the magnetostatics. But unfortunately whilst the electric point charge is one of nature's most fundamental realities, the magnetic pole is a fiction used to explain the phenomena which can be explained better by alternative means.

It was found by the physicists that it was not possible to isolate single poles, similar to the point electric charges, but that the magnetic poles existed in pairs, i.e. 'a magnetic dipole' which consists of a pair of equal and opposite magnetic 'charges' fixed together as the smallest physical unit and it could consist of (as an equivalent) an electric current in a small loop circuit. In the unmagnetized iron, the dipoles lie at random, and hence as a consequence have no

resultant external effect. In the magnetized iron, these dipoles are aligned in the direction of magnetization, so that they neutralize in the centre of the magnet, but leave the appearance of the positive and the negative charges at the ends (i.e. the north and the south polarity at the ends).

7.2 THE LAW OF MAGNETIC FORCE BETWEEN TWO SMALL MOVING CHARGES

Though a direct experimental verification of the calculated force between the two moving point charges is rather difficult, it is possible to determine that force indirectly. For example, it is possible to measure the force between the two current-carrying conductors and then reduce it mathematically to find the elemental force between the two elemental electric charges. Alternatively, the motion of a single charge in the vicinity of a current-carrying conductor can be analyzed, and then express the total force on the particle as a sum of the magnetic forces exerted on it by the individual electrons moving in the conductor.

So now we consider the two point charges Q_1 and Q_2 moving in a vacuum with the velocities $\mathbf{v}_1$ and $\mathbf{v}_2$, respectively, with respect to an observer (Figure 7.1). Let us also suppose that $\mathbf{v}_1$ and $\mathbf{v}_2$ are << c, the velocity of light.

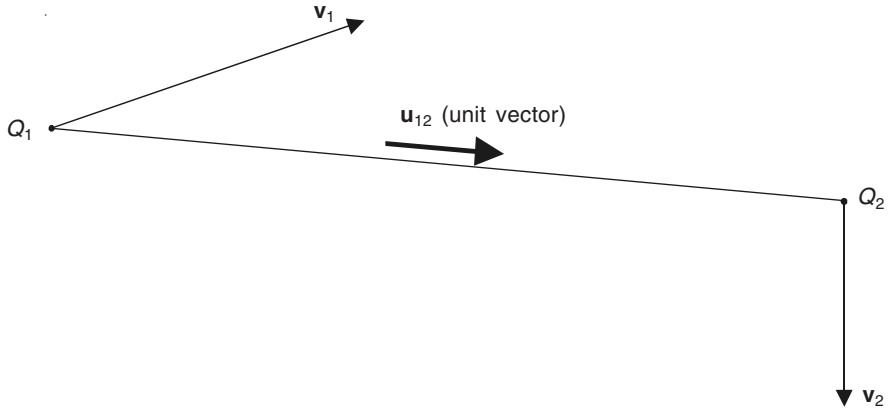


Figure 7.1 Two moving point charges.

The magnetic force $\mathbf{F}_{m12}$ exerted by the charge Q_1 on the charge Q_2 is

$$\mathbf{F}_{m12} = K_m \left(\frac{Q_1 Q_2}{r^2} \right) (\mathbf{v}_2 \times (\mathbf{v}_1 \times \mathbf{u}_{12}) \tag{7.1}$$

where K_m depends on the property of the space.

(In SI units, $K_m = \mu_0/4\pi$, where μ_0 is the 'absolute permeability of the free space'.)

It should be noted that in Eq. (7.1) the vector given by the cross product $(\mathbf{v}_1 \times \mathbf{u}_{12})$ should be evaluated first. This would then be multiplied vectorially by the velocity vector $\mathbf{v}_2$.

The unit of velocity in SI units is metres/second, and one coulomb per second is one ampere. Hence the unit of μ_0 is newtons/(ampere)^2 which can be written as henry/metre, its notation being H/m. This last equivalence will be proved later.

Comparing Eq. (7.1) for the force between the moving charges with the Coulomb's law for the force between the static charges [Eq. (1.4)], i.e.

$$F_{12} = \left(\frac{1}{4\pi\varepsilon_0}\right)\left(\frac{Q_1 Q_2}{r^2}\right) \tag{1.4}$$

we find that the product $(\mu_0\varepsilon_0)^{-1}$ has the dimension (velocity)^2.

Now we know, $\varepsilon_0 = 8.854 \times 10^{-12} = \dfrac{1}{36\pi \times 10^9}$

and the value of μ_0 is exactly given by

$$\mu_0 = 4\pi \times 10^{-7} \text{ H/m} \tag{7.2}$$

(air also has practically the same permeability).
Hence,

$$(\mu_0\varepsilon_0)^{-1} \simeq 9 \times 10^{16} = (3 \times 10^8)^2$$

or

$$\frac{1}{\sqrt{\mu_0\varepsilon_0}} = 3 \times 10^8 \text{ m/s}$$

$$\simeq c_0 \text{ (the velocity of light in vacuum)} \tag{7.3}$$

It should be noted that the magnitude of the magnetic force $\mathbf{F}_{m12}$ given by Eq. (7.1) is a function of both the magnitude and the direction of the velocities of the two charges. The largest possible magnitude of this force would be

$$|\mathbf{F}_{m12}| = \left(\frac{\mu_0}{4\pi}\right)\left(\frac{Q_1 Q_2}{r^2}\right)(v_1 v_2) \tag{7.4}$$

The ratio of the maximal possible magnetic force and the electric force is

$$\frac{|\mathbf{F}_{m12}|}{|\mathbf{F}_{e12}|} = \varepsilon_0\mu_0 v_1 v_2 = \frac{v_1 v_2}{c_0^2} \tag{7.5}$$

Since the velocities of the electric charges in practice are much smaller than the velocity of light in vacuum, the magnetic force between the charges is negligible compared with the coulomb force between them. Next, we consider the direction of the magnetic force exerted by the charge Q_2 on the charge Q_1 which is obtained by interchanging the subscripts 1 and 2 in Eq. (7.1),

$$\mathbf{F}_{m21} = \left(\frac{\mu_0}{4\pi}\right)\left(\frac{Q_1 Q_2}{r^2}\right)\left[\mathbf{v}_1 \times \left(\mathbf{v}_2 \times \mathbf{u}_{21}\right)\right] \tag{7.6}$$

Although $\mathbf{u}_{21} = -\mathbf{u}_{12}$, in general, $\mathbf{F}_{m21} \neq \mathbf{F}_{m12}$, because

$$[\mathbf{v}_2 \times (\mathbf{v}_1 \times \mathbf{u}_{12})] \neq -[\mathbf{v}_1 \times (\mathbf{v}_2 \times \mathbf{u}_{21})] \tag{7.7}$$

For example, if $\mathbf{v}_1$ is parallel to $\mathbf{u}_{12}$, then the left-hand side of Eq. (7.7) is = 0, because $(\mathbf{v}_1 \times \mathbf{u}_{12}) = 0$, but the right-hand side is not necessarily zero.

This may appear as an apparent contradiction of the Newton's third law. At this stage, it is not possible to go in depth into the understanding of this problem. But it can be explained by saying that a system of two isolated moving bodies cannot be considered as a 'two-body' problem, but has to be treated as a 'three-body' problem. Since the charges are moving all the time, their electric field varies in time at all the points, and this variation is dependent on certain

inertial properties. In a three-body system, Newton's law cannot be applied in isolation to only two of the bodies of the system.

7.3 THE CONCEPT OF THE MAGNETIC FIELD (THE MAGNETIC FLUX DENSITY)

In the expression for $\mathbf{F}_{m12}$, given by Eq. (7.1), if instead of Q_2, there is a charge Q moving with the velocity $\mathbf{v}$, then the force $\mathbf{F}_m$ can be written as

$$\begin{aligned}\mathbf{F}_m &= \left(\frac{\mu_0}{4\pi}\right)\left(\frac{Q_1 Q_2}{r^2}\right)\left[\mathbf{v}_2 \times (\mathbf{v}_1 \times \mathbf{u}_{12})\right] \\ &= Q_2\mathbf{v}_2 \times \left(\frac{\mu_0}{4\pi}\right)\left(\frac{Q_1 (\mathbf{v}_1 \times \mathbf{u}_{12})}{r^2}\right) \\ &= Q\mathbf{v} \times \left(\frac{\mu_0}{4\pi}\right)\left(\frac{Q_1\mathbf{v}_1 \times \mathbf{u}_{12}}{r^2}\right) \\ &= Q\mathbf{v} \times \mathbf{B} \qquad (7.8)\end{aligned}$$

where

$$\mathbf{B} = \left(\frac{\mu_0}{4\pi}\right)\left(\frac{Q_1\mathbf{v}_1 \times \mathbf{u}_{12}}{r^2}\right) \qquad (7.9)$$

This is the definition of the magnetic flux density vector produced by the moving charge Q. From Eq. (7.8), the unit of the magnetic flux density is newton-second per coulomb-metre. A different name tesla (notation—T) is given to this unit. We shall see later that this is also equal to webers per square metre (Wb/m^2).

Now instead of only two point charges, let us assume that there are n charges Q_1, Q_2, Q_3, ..., Q_n, moving with velocities $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$, ..., $\mathbf{v}_n$, respectively (with respect to the same coordinate system), then by the principle of superposition of the magnetic forces, the total magnetic force on Q due to these n charges will be

$$\mathbf{F}_m = Q\mathbf{v} \times \left(\frac{\mu_0}{4\pi}\right)\left(\frac{Q_1\mathbf{v}_1 \times \mathbf{u}_1}{r^2}\right) + \cdots + \frac{Q_n\mathbf{v}_n \times \mathbf{u}_n}{r^2} \qquad (7.10)$$

where $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$, . . ., $\mathbf{u}_n$, are the unit vectors directed (at the instant considered) from the charges Q_1, Q_2, Q_3, . . ., Q_n, towards the charge Q. So the magnetic flux density vector $\mathbf{B}$ due to the n moving charges is

$$\mathbf{B} = \left(\frac{\mu_0}{4\pi}\right)\sum_{k=1}^{n} \frac{Q_k\mathbf{v}_k \times \mathbf{u}_k}{r_k^2} \qquad (7.11)$$

To generalize, suppose in the region there are some static charges as well, and Q is subject to an electric field force $\mathbf{F}_e$ which will be $\mathbf{F}_e = Q\mathbf{E}$, where $\mathbf{E}$ is the electric field intensity due to the static charges.

Hence the total force on the charge Q is

$$\mathbf{F} = \mathbf{F}_e + \mathbf{F}_m = Q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \qquad (7.12)$$

This above total force is known as the 'Lorentz force'.

So far we have considered the magnetic field of a single point charge or a number of discrete point charges, and the magnetic flux density **B** created by such moving charges, changes with time at all points. Such a field is not a static field. However during the initial part of our study, we shall consider only the static magnetic fields, that is, the magnetic fields which are independent of time or do not vary with time. A macroscopically static magnetic field is produced either by a permanent magnet (which we shall study in due course), or when there is a 'stationary system of electric charges in motion' by which we mean a steady electric current whose magnitude is independent of time. In practice, the static magnetic fields are produced by the steady electric currents in metallic conductors. But it *must* be remembered that the magnetic field is present whenever there is a relative motion between the electric charges and the observer.

7.4 THE MAGNETIC FIELD OF AN ELECTRIC CURRENT—BIOT–SAVART'S LAW

From the expressions for the magnetic force due to the discrete moving electric charges, the next logical step would be to derive the expressions for the magnetic field due to steady electric current in conductors in which the steady current is produced by a continuous steady stream of free electrons. However though this would seem to be a logical sequence, the historical development of magnetism has been somewhat different. As it was mentioned earlier, the concept of the magnetic charge (or the isolated magnetic pole), parallel to the concept of the electric charges, led the development of this science along a somewhat different path. So we diverge for a while to see how the actual development took place.

It was in 1820, Hans Christian Oersted, a Danish professor, demonstrated, for the first time, to his students that an electric current produced a magnetic field in its vicinity. This was the first of the great unifying discoveries in the process of linking the 'separate' branches of the electromagnetic science. Also Oersted's discovery implied that when a current-carrying wire is placed in a magnetic field, the wire would be subjected to a mechanical force. As a consequence it follows that two current-carrying circuits must exert forces on each other. The nature of the magnetic field of a current in a circuit was explored in very great depth and detail by the French academician André Marie Ampere (1775–1836) whom Maxwell has called the 'Newton of electricity'. By a series of very clever and original experiments, Ampere, with remarkable insight, determined the forces of interaction between two circuits, and concluded that 'even the magnetism of a bar magnet could be ascribed to currents within the material of the bar.' This hypothesis agrees with the modern physical theory, because the atomic orbits with their circling electrons may be regarded as tiny loops of currents. There is also another source of atomic magnetism, i.e. the electron spin and the nuclear spin which also could be justifiably conceived as producing the tiny loops of current within the material. Henceforth we shall refer to a current flowing in a small circuit as a 'current loop'. This could be a loop of atomic dimensions or a small loop of wire or a small coil of many turns. The current can be introduced into it from a remote source by means of two wires running very close together. The magnetic field of these two wires, so close together and carrying currents in opposite directions, would be negligible. The current loop is thus a practical piece of laboratory apparatus.

So now we are in a position to consider the equivalence between a 'magnetic dipole' and a 'current loop'. The magnetic dipole would be the ultimate magnetic particle of microscopic dimensions, having the positive and the negative poles of identical strength; and we have seen above that the current loop is the ultimate source of the magnetic flux. These two approaches are

magnetically equivalent; that is, if they both existed in nature, it would not be possible to distinguish them by any magnetic experiment. This was seen experimentally first by Ampere and then by Weber by their experiments on small current loops and bar magnets. Quantitatively they found that the moment of the equivalent dipole (= the product of the pole-strength with the distance between the poles) is proportional to the product of the current with the area of the circuit. So, if a current i is flowing in a small loop of area S, then it is magnetically equivalent to a dipole of moment m, given by

$$m = \mu_0 iS \tag{7.13}$$

The dipole is directed in such a sense that a right-hand screw turned in the direction of the current in the loop would move from the negative end to the positive end of the dipole (Figure 7.2). So now it is possible to think of a small bar magnet to be equivalent to an elongated coil with current flowing round its periphery (Figure 7.3).

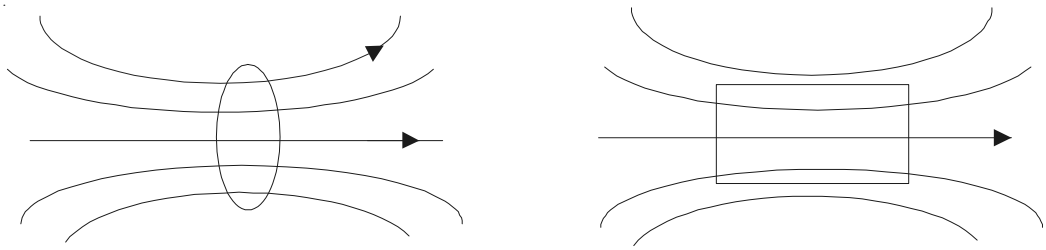


Figure 7.2 Equivalence between a current loop and a magnetic dipole.

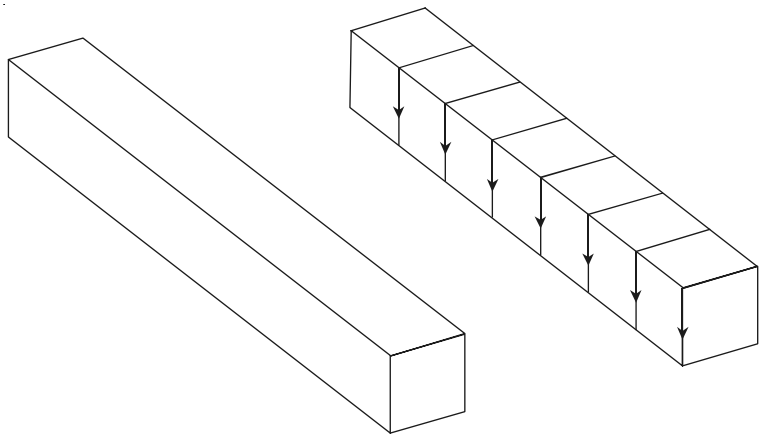


Figure 7.3 A bar magnet and its current loop equivalent.

Though our basis for the magnetism is 'the current loop', our attitude to the current loops and the dipoles should be utilitarian, and we would not hesitate to use either concept to solve practical problems, choosing the option which gives the quicker result.

With this background, we now go back to the study of the magnetic field of steady electric currents, starting from Eq. (7.10) which we rewrite as

$$\mathbf{F}_{\mathrm{m}} = Q\mathbf{v} \times \left(\frac{\mu_0}{4\pi}\right) \sum_{k=1}^{n} \frac{Q_k \mathbf{v}_k \times \mathbf{u}_k}{r_k^2}$$

The current-carrying conductor is shown in Figure 7.4. If the test charge ΔQ [the Q of Eq. (7.10) has been replaced by ΔQ] is at a fixed point outside the conductor, then it will be acted upon

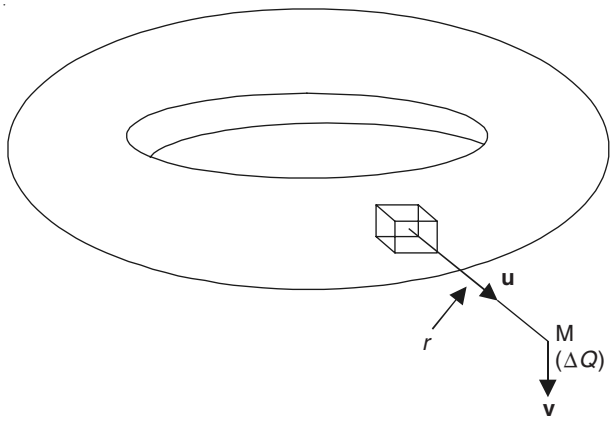


Figure 7.4 A small charge in motion in the field of charges moving inside a conductor.

neither by the electric force nor by the magnetic force. The moving charges inside the conductor are compensated, and hence there are no appreciable electric forces outside or inside the conductor. When the test charge ΔQ is stationary, the magnetic force on ΔQ is also zero, as $\mathbf{v} = 0$. When the test charge ΔQ is set in motion with a velocity $\mathbf{v}$, the electric force remains zero, but now the magnetic force on it becomes

$$\mathbf{F}_{\mathrm{m}} = \Delta Q\mathbf{v} \times \left(\frac{\mu_0}{4\pi}\right)\sum_{k=1}^{n} \frac{Q\mathbf{v}_k \times \mathbf{u}_k}{r_k^2}$$

since all the free electrons in the conductor carry the same charge Q and their average drift velocities are $\mathbf{v}_k$. So the magnetic flux density at any point is given by

$$\mathbf{B} = \left(\frac{\mu_0}{4\pi}\right)\sum \frac{Q\mathbf{v}_k \times \mathbf{u}_k}{r_k^2} \tag{7.14}$$

Let us say that the number of the free electrons per unit volume of the conductor is N. Then, the number of charges in a small volume, Δv is = $N\Delta v$. They all move with the same velocity $\mathbf{v}$, as Δv is assumed to be very small. Hence the sum of the individual charges in Eq. (7.14) would be replaced by the sum over the volume elements, i.e.

$$\mathbf{B} = \left(\frac{\mu_0}{4\pi}\right)\sum_{v}\left(\frac{NQ\mathbf{v} \times \mathbf{u}_k}{r^2}\right)\delta v \tag{7.15}$$

But according to Eq. (6.1), $NQ\mathbf{v} = \mathbf{J}$, the current density vector at the points inside the volume element δv. If δv is small enough, then the summation sign can be replaced by a volume integral. Hence the magnetic flux density due to the steady current in the conductor is given by

$$\mathbf{B} = \left(\frac{\mu_0}{4\pi}\right)\iiint_{v}\left(\frac{\mathbf{J} \times \mathbf{u}}{r^2}\right)dv \tag{7.16}$$

where $\mathbf{u}$ is the unit vector directed from the volume element dv towards the fixed point at which the flux density is being determined. In practice, the current is often flowing through thin wires.

Let us assume the cross-section of the wire to be ΔS. Then, $\delta v = \Delta S dl$, where dl represents the length element of the wire. Now $J\Delta S$ ($= \mathbf{J} \cdot \Delta \mathbf{S}$) is the current intensity or the current I which is not a vector quantity. So it is conventional to orient dl in the direction of $\mathbf{J}$. Hence the magnetic flux density vector of a thin wire carrying a current I is

$$\mathbf{B} = \left(\frac{\mu_0 I}{4\pi} \right) \int_C \left(\frac{d\mathbf{l} \times \mathbf{u}}{r^2} \right) \tag{7.17}$$

and for every element of the wire, this equation is

$$\Delta \mathbf{B} = \left(\frac{\mu_0 I}{4\pi} \right) \left(\frac{d\mathbf{l} \times \mathbf{u}}{r^2} \right) \tag{7.18}$$

(See Figure 7.5.)

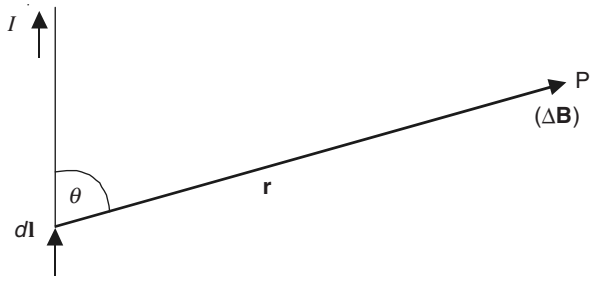


Figure 7.5 Magnetic field due to a current element.

Equation (7.18) in this form is known as Biot–Savart's law. It should again be noted that though we have derived this equation in a logical manner starting from the point charges in motion, its discovery was made by two French scientists Biot and Savart based on their experimentation and some very clever mathematical insight. Note that the direction of $\Delta \mathbf{B}$ in Figure 7.5, is normal to the plane containing dl and r, i.e. in the circumferential direction normal to the plane of the paper. The magnitude of the flux density is

$$\left| \Delta \mathbf{B} \right| = \frac{\mu_0 I dl \sin \theta}{4\pi r^2} \tag{7.19}$$

We shall now illustrate the application of the Biot–Savart's law for evaluating the magnetic flux density vector $\mathbf{B}$ in some simple geometrical configurations.

7.4.1 Magnetic Field of a Short Straight Length of Wire

This is a conceptual problem, because in reality it is not possible to have an isolated piece of straight current-carrying conductor of finite length. But this result can be used to solve many real practical problems.

We consider a finite length of wire (assumed to be infinitely thin), carrying a current I. We wish to find the magnetic field at a point P (Figure 7.6). So we take an element length δl, whose distance from P is r, as shown, which length makes an angle θ with the line of the conductor. Hence,

$$\mathbf{B}_P = \mu_0 \int_{\substack{\text{over the length of} \\ \text{the conductor}}} \frac{I \sin \theta \, dl}{4\pi r^2} \tag{7.20}$$

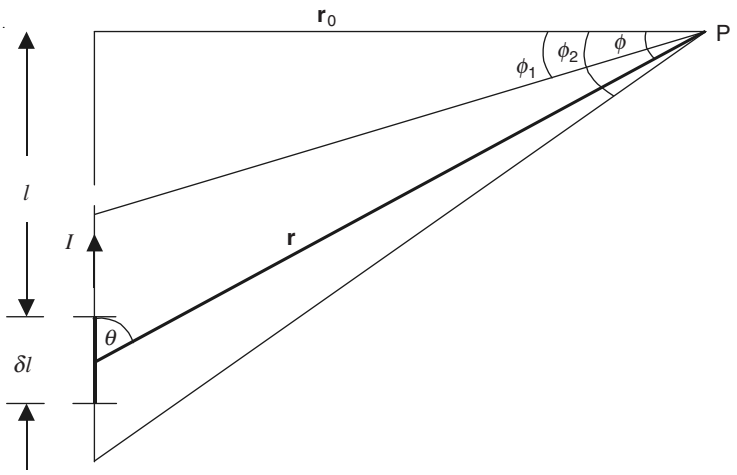


Figure 7.6 Finite length of a current-carrying conductor.

In the preceding expression, there are three apparently independent variables—θ, l, and r—which in fact are interrelated. So we convert them to a single independent variable ϕ which as the Figure 7.6 shows, is the complement to the angle θ. So the length of the conductor can also be expressed in terms of ϕ, i.e. the top and the bottom ends of the conductor make the angles ϕ_1 and ϕ_2 with the perpendicular from P on to the line of the conductor. The length of the perpendicular is r_0 and all the lengths along the conductor are measured from the foot of the perpendicular. So we have the relationships

$$l = r_0 \tan \phi, \quad \therefore \quad dl = r_0 \sec^2 \phi \, d\phi$$

Also

$$r = r_0 \sec \phi \qquad \text{and} \qquad \sin \theta = \cos \phi$$

$\therefore$ For the length of the conductor, the magnetic field **B** is

$$B_P = \mu_0 \int_{\phi=\phi_1}^{\phi=\phi_2} \left(\frac{I}{4\pi}\right)\left(\frac{\cos\phi \, r_0 \sec^2\phi \, d\phi}{r_0^2 \sec^2\phi}\right)$$

$$= \frac{\mu_0 I}{4\pi r_0} \int_{\phi_1}^{\phi_2} \cos\phi \, d\phi$$

$$= \frac{\mu_0 I}{4\pi r_0} (\sin\phi_2 - \sin\phi_1) \tag{7.21}$$

For the direction of the current shown in the figure, the direction of $\mathbf{B}_P$ would be normal to the plane containing dl and r, i.e. the plane of the paper in this case, and it is directed 'into' the paper. To find the magnetic field **B** due to an infinitely long wire carrying a current I, we have to substitute $\phi_1 = -\pi/2$, $\phi_2 = +\pi/2$ in Eq. (7.21). So, in this case

$$B_P = \frac{\mu_0 I}{2\pi r_0} \text{ tesla} \tag{7.22}$$

7.4.2 Magnetic Field on the Axis of a Square Coil

ABCD is the square coil (Figure 7.7) of side a, carrying a current I, and OP is the axis of the coil, where the height of the point P from the plane of the square is b, i.e. OP = b. In the construction

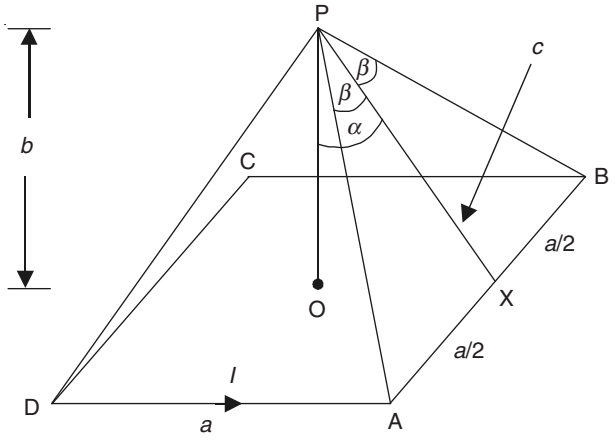


Figure 7.7 A square coil.

shown, X is the mid-point of the side AB, and obviously, $\angle$ APX = $\angle$ BPX = β. The magnitude of the magnetic flux density due to the current I in each side of the square will be the same though its direction will be different, for example, let us consider the side AB. The magnetic flux density due to the side AB will have the magnitude

$$B_S = \left(\frac{\mu_0 I}{4\pi c}\right)\left[\sin\beta - \sin(-\beta)\right] = \left(\frac{\mu_0 I}{2\pi c}\right)\sin\beta \qquad \text{[from Eq. (7.21)]}$$

where

$$c = \sqrt{b^2 + (a/2)^2}$$

$$\sin\beta = \frac{a/2}{\sqrt{(a/2)^2 + c^2}} = \frac{a}{\sqrt{a^2 + 4c^2}} = \frac{a}{\sqrt{2a^2 + 4b^2}}$$

The direction of B_S due to the current in AB will be normal to the plane containing the ΔABP. Because of the contributions from all the four sides, the horizontal components of B_S due to the two sets of parallel sides will cancel out, leaving only the vertical components.

$\therefore$

$$\begin{aligned} B_{TV} &= 4\left(\frac{\mu_0 I}{2\pi c}\right)\sin\beta\sin\alpha \\ &= \left(\frac{2\mu_0 I}{\pi c}\right)\left(\frac{a}{\sqrt{2a^2 + 4b^2}}\right)\left(\frac{a/2}{c}\right) \\ &= \frac{\mu_0 I a^2}{\pi c^2\sqrt{2a^2 + 4b^2}} \\ &= \frac{4\mu_0 I a^2}{\pi\left(a^2 + 4b^2\right)\sqrt{2a^2 + 4b^2}} \end{aligned} \qquad (7.23)$$

7.4.3 Magnetic Field on the Axis of a Circular Coil

Figure 7.8 shows a circular coil of radius a, carrying a current I. The point P under consideration is on the axis of the coil at a height b from the plane of the coil. The magnetic field $\delta\mathbf{B}$ at P due to a circumferential element δl as shown in Figure 7.8, is

$$\delta\mathbf{B} = \frac{\mu_0 I \delta l \sin(\pi/2)}{4\pi c^2} = \frac{\mu_0 I \delta l}{4\pi c^2}$$

The direction of $\delta\mathbf{B}$ is such that it makes an angle $(\theta + \pi/2)$ to the axis OP. It should be noted that θ of Eq. (7.19), is equal to $\pi/2$ in this problem as marked in Figure 7.8. Now, there is an

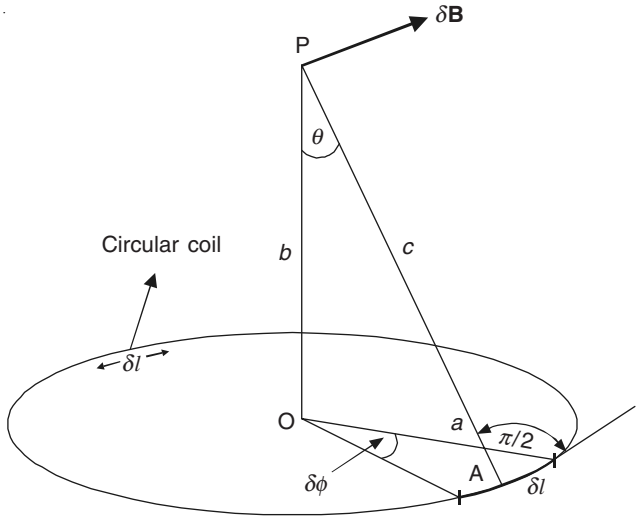


Figure 7.8 A circular coil.

equal and opposite element $I\delta l$, diametrically opposite to the element under consideration, as shown in the figure. So by symmetry, all the components of $\mathbf{B}$ normal to the axis OP will cancel out; and the components along the axis add up, i.e.

$$B_{ax} = \frac{\mu_0}{4\pi}\int_{\phi=0}^{\phi=2\pi}\left(\frac{I\delta l}{c^2}\right)\cos(\pi/2 - \theta) \quad \text{and} \quad \delta l = a\delta\phi \quad \text{and} \quad \sin\theta = a/c$$

$$= \frac{\mu_0 I}{4\pi}\int_{\phi=0}^{\phi=2\pi} ad\phi\left(\frac{\sin\theta}{c^2}\right) = \left(\frac{\mu_0 I \sin^3\theta}{4\pi a}\right)[\phi]_0^{2\pi}$$

$$= \left(\frac{\mu_0 I}{2a}\right)\sin^3\theta \tag{7.24}$$

When the point P is at the centre of the circle, i.e. in the plane of the circle, $\theta = \pi/2$.

$$\therefore \quad B \text{ at O} = \frac{\mu_0 I}{2a} \tag{7.25}$$

7.4.4 Magnetic Field on the Axis of a Short Circular Solenoid

Let us consider a circular solenoid of finite length (axial), as shown in Figure 7.9, which has n turns per metre, each turn carrying a current I. We wish to find the magnetic field at a point P on

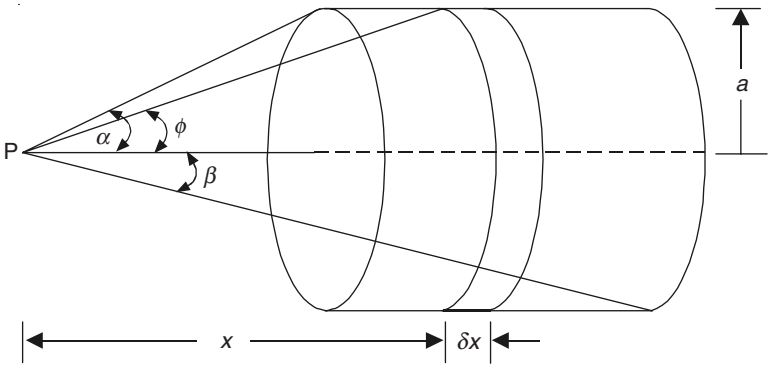


Figure 7.9 A circular solenoid of finite axial length.

the axis of the solenoid, where P subtends semiangles α and β at the near end and the far end of the solenoid, respectively. We take an elemental strip of the solenoid of axial width δx where the axial distance of the strip from the point P is x. We now use the result of the last problem [Eq. (7.24)] to solve this problem.

The current flowing in the elemental strip of the solenoid = $nI\delta x$. Expressing x in terms of the semiangle ϕ subtended by the strip to the point P, we have, $x = a \cot \phi$, where a is the radius of the solenoid.

$\therefore$ $$\delta x = -a\,\text{cosec}^2\phi\,\delta\phi$$

$\therefore$ From Eq. (7.24),

$$\delta B = \left(\frac{\mu_0 nI\delta x}{2a}\right)\sin^3\phi = -\left(\frac{\mu_0 nIa}{2a}\right)\text{cosec}^2\phi\sin^3\phi\,\delta\phi$$

$$= -\left(\frac{\mu_0 nI}{2}\right)\sin\phi\,\delta\phi$$

$\therefore$ Due to the whole solenoid,

$$B = -\left(\frac{\mu_0 nI}{2}\right)\int_{\phi=\alpha}^{\phi=\beta}\sin\phi\,\delta\phi$$

$$= \left(\frac{\mu_0 nI}{2}\right)(\cos\beta - \cos\alpha) \tag{7.26}$$

If now the solenoid is infinitely long, then $\beta = 0$ and $\alpha = \pi$, and hence

$$B = \mu_0 nI \tag{7.27}$$

7.4.5 Magnetic Flux Density of Planar Currents, at an Arbitrary Point in Their Plane

Let us consider an arbitrary contour lying in a plane as shown in Figure 7.10. The magnetic flux

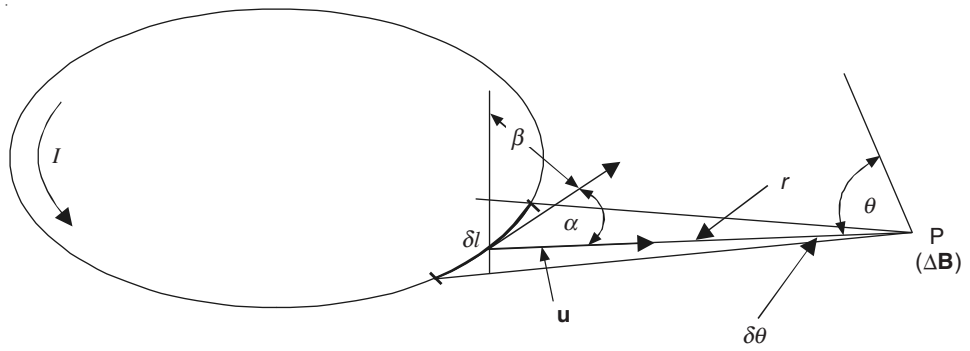


Figure 7.10 A planar current-carrying contour.

density at the point P lying in the same plane as the circuit is required. We take an element δl of the contour as shown in the figure. Then using Eq. (7.18) of the Biot–Savart's law,

$$\Delta \mathbf{B} = \left(\frac{\mu_0 I}{4\pi}\right)\left(\frac{\delta \mathbf{l} \times \mathbf{u}}{r^2}\right) \tag{7.18}$$

We have for the element shown in Figure 7.10:

$$|(\delta \mathbf{l} \times \mathbf{u})| = \delta l \sin \alpha \qquad \text{(and is normal to the plane of the paper)}$$

and

$$\delta l = \frac{r\,\delta\theta}{\cos\beta} = \frac{r\,\delta\theta}{\cos\left(\frac{\pi}{2} - \alpha\right)} = \frac{r\,\delta\theta}{\sin\alpha}$$

$$\therefore \qquad \Delta B = \left(\frac{\mu_0 I}{4\pi}\right)\left(\frac{\delta\theta}{r}\right) \tag{7.28}$$

So the magnetic field at P due to the complete closed circuit would be

$$B = \left(\frac{\mu_0 I}{4\pi}\right)\oint_C \frac{d\theta}{r} \tag{7.29}$$

The direction of B would be normal to the plane of the paper and into it.

7.5 THE LINES OF MAGNETIC FLUX DENSITY VECTOR B, AND THE MAGNETIC FLUX Φ

We define the lines of the magnetic flux density vector **B** as those lines for which the vector **B** is tangential at all the points. When we draw a set of such lines for a magnetic field and mark them by arrows indicating the direction of the vector **B**, then the direction of **B** is known at all such points. These lines are very useful for visualizing the magnetic field in a region.

From the example of the Section 7.4.1, and Eq. (7.22), it is obvious that the lines of **B** for an infinitely long straight current filament, are concentric circles with their centres at the axis of the filament (Figure 7.11).

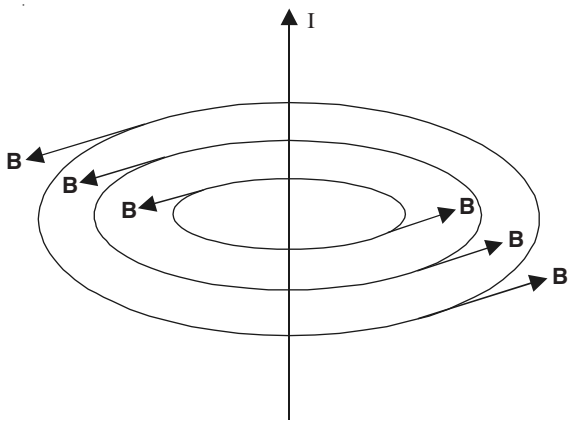


Figure 7.11 Lines of **B** for a straight current filament.

Next, we consider the flux of the vector **B**, i.e. the flux of the magnetic flux density vector **B** through a surface *S*. It is called the magnetic flux which has the notation Φ (Figure 7.12).

$$\Phi = \oiint_S \mathbf{B} \cdot d\mathbf{S} \tag{7.30}$$

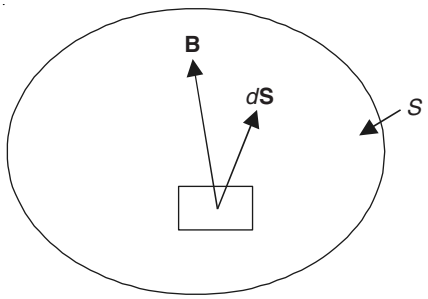


Figure 7.12 Surface defining the magnetic flux.

The unit of the flux is webers, the notation being Wb. So the unit of **B** which is tesla (= T), is now T = Wb/m^2. The tubular surfaces formed by the lines of the vector **B** are defined as the 'tubes of magnetic flux'. So the vector **B** is tangential to such a tube at all points. Hence furthemore the magnetic flux through any cross-section of a tube is constant.

7.6 THE LAW OF CONSERVATION OF MAGNETIC FLUX

The magnetic flux has a simple and yet an extremely important property: *the magnetic flux through any closed surface is zero*. This is also known as the *law of conservation of magnetic flux*. It is proved as follows:

The elemental charges in motion produce a magnetic field.

$\therefore$ The total **B** in a field = Sum of the elemental **B** vectors caused by all the elemental charges in a system.

Also, according to Eq. (7.30):

The magnetic flux Φ through any surface = Sum of the elemental Φs through that surface, due to all the elemental moving charges which produce the magnetic field.

Let us consider a closed surface S situated in the magnetic field of the system. Initially let us assume the system to be due to a single moving charge. The lines of **B** would be concentric circles with their centres on the line of motion of the charge. So any line of **B** would enter the closed surface exactly as many times as it (the line of **B**) leaves the surface. The lines represent tubes of equal flux.

$\therefore$ Inward flux of **B** = Outward flux of **B**

through the closed surface S:

$\therefore$ The magnetic flux, through any closed surface, produced by any number of moving charges is zero.

$\therefore$ The law of conservation of magnetic flux (in integral form) is

$$\oiint_S \mathbf{B} \cdot d\mathbf{S} = 0 \tag{7.31}$$

Note that by Gauss' divergence theorem, we have for any vector **A**

$$\oiint_S \mathbf{A} \cdot d\mathbf{S} = \iiint_v \text{div } \mathbf{A}\, dv$$

where v is the volume enclosed by the closed surface S.

$\therefore$ Equation (7.31) becomes

$$\iiint_v \text{div } \mathbf{B}\, dv = 0$$

$\therefore$
$$\text{div } \mathbf{B} = \nabla \cdot \mathbf{B} = 0 \tag{7.32}$$

This is the differential form of the law of conservation of magnetic flux, and one of the Maxwell's equations of electromagnetic fields.

A mathematical proof of Eq. (7.32)

We have, from Eq. (7.16), the expression for **B** due to a steady current in the conductor as

$$\mathbf{B} = \left(\frac{\mu_0}{4\pi}\right) \iiint_v \frac{\mathbf{J} \times \mathbf{u}}{r^2}\, dv \tag{7.16}$$

Consider the divergence of **B** at a point P

$$\text{div } \mathbf{B} = \left(\frac{\mu_0}{4\pi}\right) \nabla_P \cdot \left[\iiint_v \frac{\mathbf{J} \times \mathbf{u}}{r^2}\, dv\right]$$

The unit vector **u** is directed from the volume element dv to the field point P. (Note that the volume element and the integral refer to the volume occupied by the currents.)

$$\therefore \quad \frac{\mathbf{u}}{r^2} = -\text{ grad } \frac{1}{r} = -\nabla \frac{1}{r}$$

$$\therefore \qquad \nabla \cdot \mathbf{B} = -\left(\frac{\mu_0}{4\pi}\right) \iiint_v \nabla_P \cdot \left[\mathbf{J} \times \nabla\left(\frac{1}{r}\right)\right] dv \tag{7.33}$$

is obtained by interchanging the operations 'div_P' and the volume integral, which are independent operations (and hence are interchangeable in their orders of performance).

Now, from Eq. (0.66)

$$\nabla \cdot (\mathbf{A} \times \mathbf{B}) = \mathbf{B} \cdot (\nabla \times \mathbf{A}) - \mathbf{A} \cdot (\nabla \times \mathbf{B})$$

$\therefore$ Equation (7.33) becomes

$$\nabla \cdot \mathbf{B} = -\left(\frac{\mu_0}{4\pi}\right) \iiint_v \left[\nabla_P\left(\frac{1}{r}\right) \cdot (\nabla_P \times \mathbf{J}) - \mathbf{J} \cdot \nabla_P \times \nabla_P\left(\frac{1}{r}\right)\right] dv$$

The second term above is (curl grad) and is identically zero. In the first term (curl_P **J**) relates to the point P and **J** is a function of the coordinates of the volume δv and so (curl_P) must be zero.

$$\therefore \qquad \nabla \cdot \mathbf{B} = 0$$

7.7 AMPERE'S LAW

We start from the results derived from the application of the Biot–Savart's law to find that the magnetic field **B** at a point P distant r from an infinitely long straight conductor carrying a current I, is (Figure 7.13) given by

$$B = \frac{\mu_0 I}{2\pi r}$$

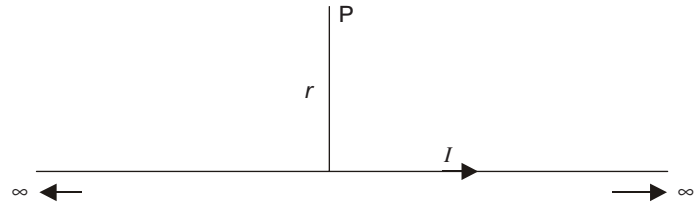


Figure 7.13 An infinitely long conductor carrying current I.

The direction of **B** being circumferential, i.e. normal to the plane of the paper, and coming out of it, we have

$$B \times (2\pi r) = \mu_0 I \tag{7.34}$$

i.e. the vector **B** multiplied by the length of the contour (over which B is constant, and also follows the direction of **B**), is proportional to the current in the wire.

At this stage, we define a new vector **H** such that

$$\mathbf{H} = \frac{\mathbf{B}}{\mu}\left(= \frac{\mathbf{B}}{\mu_0} \text{ in free space}\right) \tag{7.35}$$

so that Eq. (7.34) becomes

$$H \times (2\pi r) = I \tag{7.36}$$

Note: This **H** is an auxiliary vector which is useful in relating the **B** field to its source current I, similar to the one in electrostatics where we used $\mathbf{D}\,(=\varepsilon\mathbf{E})$ as an auxiliary vector useful in relating the **E** vector to its source charge Q in Gauss' theorem.

Generalization of the contour of **H**. Up to this stage the contour of **H** has been a circle lying in a single plane normal to the current-carrying conductor. We now consider a contour of **H** consisting of two arcs of circles both lying in the same plane, but of different radii with radial connections as shown in the Figure 7.14.

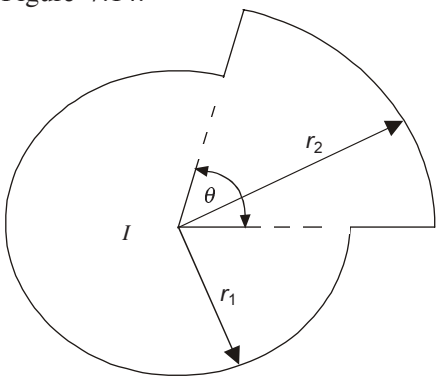


Figure 7.14 First step of the generalized contour of $\mathbf{H}\cdot d\mathbf{l}$ in the plane normal to the current filament.

Now, multiplying the component of **H** along the contour by the length of the contour, we have

$$\left(\frac{I}{2\pi r_2}\right)\times r_2\theta+\left(\frac{I}{2\pi r_1}\right)(2\pi-\theta)r_1=I \tag{7.37}$$

The contributions along the radial paths of length $(r_2 - r_1)$ will be zero as all **H** is circumferential.

From this step it is obvious that a contour of any shape can be built up of infinite number of circumferential elements and radial steps, and the expression (7.37) can be generalized for a contour of any arbitrary shape as

$$\oint_C \mathbf{H}\cdot d\mathbf{l}=I \tag{7.38}$$

where I is the current enclosed by the contour C.

Next, we consider another contour made up as shown in Figure 7.15, and take the line integral of **H** around this contour as before.

In this case

$$\left(\frac{I}{2\pi r_1}\right)\times r_1\theta-\left(\frac{I}{2\pi r_2}\right)\times r_2\theta=0$$

As before, any arbitrary shaped contour can be built up of radial and circumferential elements. So

$$\oint_C \mathbf{H}\cdot d\mathbf{l}=0 \tag{7.39}$$

when the contour C does not enclose any current.

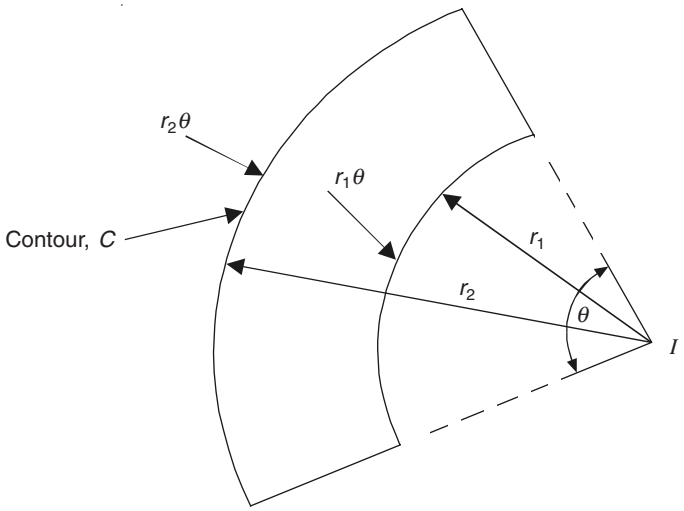


Figure 7.15 A contour not enclosing the current filament.

The contour C can be generalized still further and it does not have to lie in a single plane as shown in Figure 7.16.

$$\therefore \qquad \oint_C \mathbf{H} \cdot d\mathbf{l} = I \text{ or } 0 \tag{7.40}$$

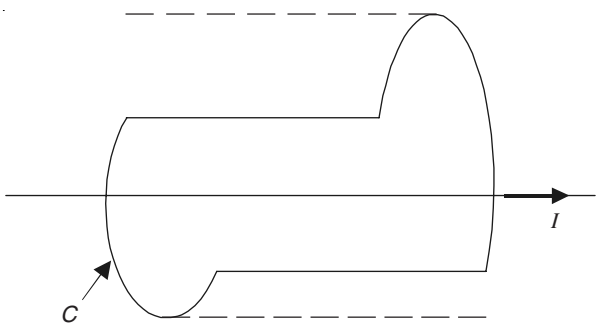


Figure 7.16 A non-coplanar contour for ($\mathbf{H} \cdot d\mathbf{l}$).

depending on whether the contour C encloses the current I or not, is true for any arbitrary contour of any shape and orientation.

Since we can also conclude that

$$\oint_C \mathbf{H} \cdot d\mathbf{l} = 0$$

for a non-linking contour, we can introduce a non-linking current-carrying return wire (also infinitely long), without altering

$$\oint_C \mathbf{H} \cdot d\mathbf{l} = I$$

for a contour linking an infinitely long current-carrying wire, as shown in Figure 7.17.

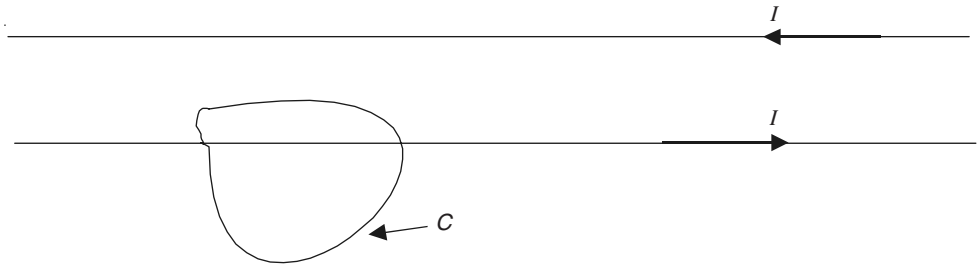


Figure 7.17 Effect of a return wire on the $\mathbf{H} \cdot d\mathbf{l}$ contour.

So far we have generalized the shape of the contour C, but considered only the infinitely long current-carrying filaments. But, since we find that the presence of non-linking wires in the vicinity of the contour has no effect on the line integral of **H,** we can generalize further and consider a circuit of any shape which can be built out of infinitely long current-carrying wires, as shown in Figure 7.18. The figure shows the building of a rectangular circuit, and so by

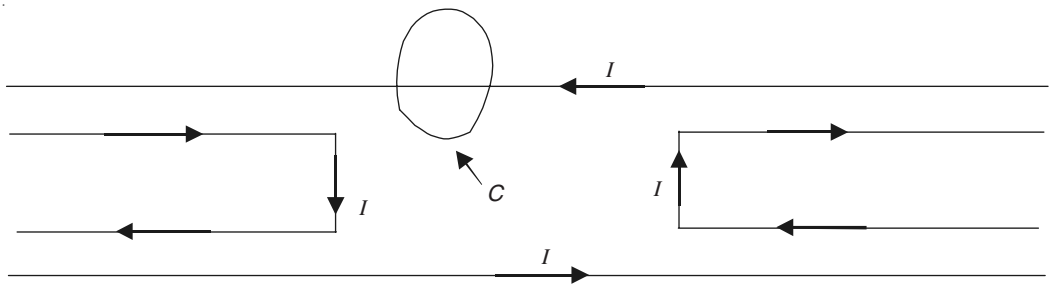


Figure 7.18 A rectangular circuit made from infinitely long current-carrying wires.

introducing as many other non-linking wires as we please, we can construct *a closed circuit of any arbitrary shape*.

In this derivation, till the present stage, we have considered only the current filaments. By using the principle of superposition, we now extend the law to conductors of finite cross-section and hence to any conducting medium having a current distribution in it.

$$\therefore \qquad \oint_C \mathbf{H} \cdot d\mathbf{l} = \Sigma I \tag{7.41}$$

Thus we have expressed the Biot–Savart's law in the form of Ampere's law (also known as the Magnetic circuit law).

If in the closed contour, C is enclosed by the current-carrying region whose cross-sectional area normal to the direction of the current is A and if the total current (ΣI) is distributed in this area, then the current density vector **J** is given as

$$|\mathbf{J}| = \frac{\Sigma I}{A} \tag{7.42}$$

$\therefore$ Substituting in Eq. (7.41), we get

$$\frac{\oint_C \mathbf{H} \cdot d\mathbf{l}}{A} = J \tag{7.43}$$

In the limit, as the contour C becomes smaller and smaller and $A \to 0$, then

$$\text{curl } \mathbf{H} = \nabla \times \mathbf{H} = \mathbf{J} \tag{7.44}$$

This is the restricted form of one of the Maxwell's equations.

7.8 MAGNETIC SCALAR POTENTIAL

If, in Eq. (7.44), the contour does not enclose any current, then

$$\text{curl } \mathbf{H} = \nabla \times \mathbf{H} = 0$$

or

$$\oint_C \mathbf{H} \cdot d\mathbf{l} = 0 \tag{7.45}$$

Then the vector $\mathbf{H}$ can be expressed as

$$\mathbf{H} = -\text{ grad } \Omega = -\nabla\Omega \tag{7.46}$$

where Ω is called the 'Scalar magnetic potential'.

The above equation holds, because

$$\oint_C (\text{grad } \Omega) \cdot d\mathbf{l} = 0 \tag{7.47}$$

is an identity.

If instead of traversing the complete closed contour, we traverse a part of the contour, say, from P to Q (the points on the contour C), then

$$\Omega_Q - \Omega_p = -\int_P^Q \mathbf{H} \cdot d\mathbf{l} \tag{7.48}$$

The analogy of the above equation with the scalar electric potential should be noted. But the magnetic scalar potential is restricted to the regions containing no currents. So, if we have a region and a contour as shown in Figure 7.19, then depending on the path chosen, we have

$$\int_P^Q \mathbf{H} \cdot d\mathbf{l} \text{ via the point R} = I$$

$$\int_P^Q \mathbf{H} \cdot d\mathbf{l} \text{ via the short route} = 0, \text{ and}$$

$$\int_P^Q \mathbf{H} \cdot d\mathbf{l} \text{ going round the path twice} = 2I \tag{7.49}$$

So it is seen that

$$\oint_C \mathbf{H} \cdot d\mathbf{l}$$

is not a single-valued function, whereas Ω is a single-valued function. Thus the concept of the

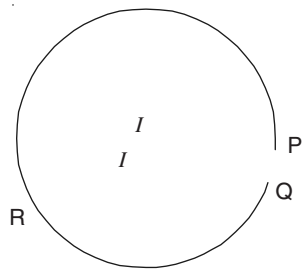


Figure 7.19 Multivalued **H**.

magnetic scalar potential can be used only if the contours are such that they do not enclose any current-carrying regions. This is achieved by the use of hypothetical 'magnetic barriers' which prevent the crossing by the contours and hence permit us to use Ω to solve such problems. This is explained in Figure 7.20.

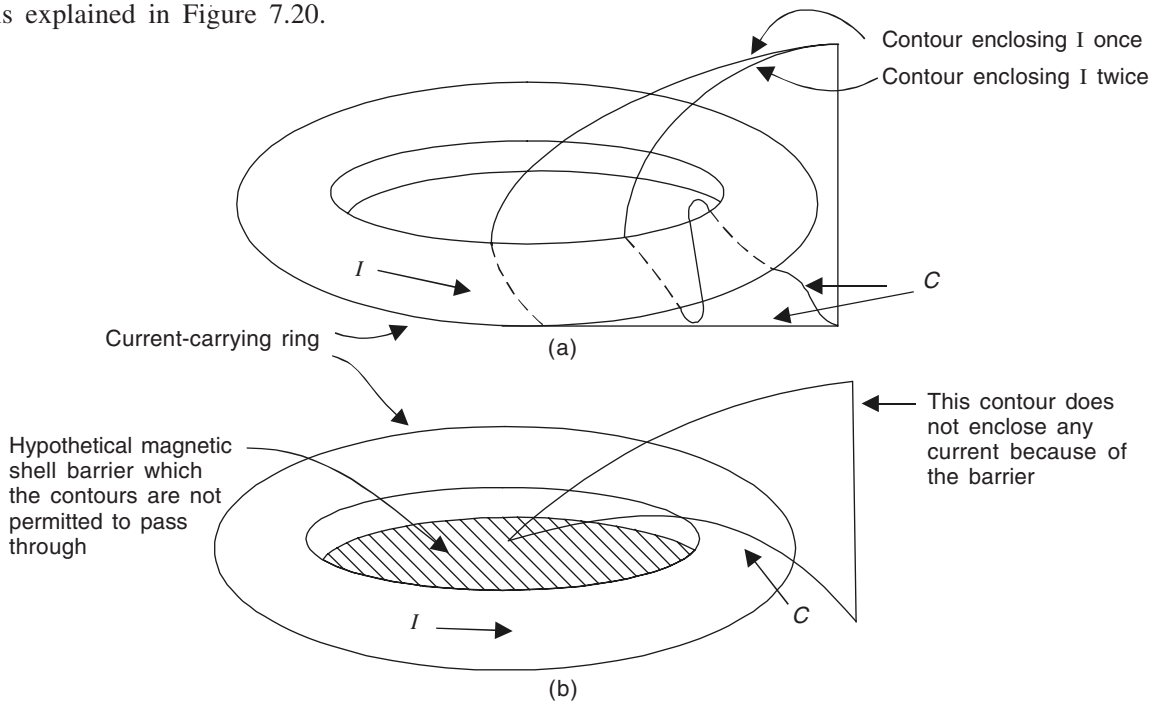


Figure 7.20 (a) Multivaluedness of ($\mathbf{H} \cdot d\mathbf{l}$) and Ω; (b) magnetic barrier to avoid the multivalued nature of ($\mathbf{H} \cdot d\mathbf{l}$) and Ω.

The unit of H and Ω. Equation (7.36) shows that the unit of **H** is ampere-turns per metre (= A/m), and the unit of Ω is ampere (= A) or ampere-turn.

Also, from the law of conservation of magnetic flux,

$$\text{div } \mathbf{B} = \nabla \cdot \mathbf{B} = 0$$

and

$$\mathbf{B} = \mu\mathbf{H} = \mu(-\text{ grad } \Omega)$$

$\therefore$ $$\text{div } [\mu(-\text{ grad } \Omega)] = -\mu \text{ div grad } \Omega$$

$$= -\mu \nabla \cdot (\nabla\Omega) = -\mu \nabla^2 \Omega = 0$$

$$\therefore \qquad \nabla^2 \Omega = 0 \tag{7.50}$$

$\therefore$ The magnetic scalar potential satisfies the Laplace's equation.

7.8.1 Scalar Magnetic Potential at a Point, due to a Current Loop of any Shape

For a large circuit as shown in Figure 7.21(a), the leads will have no magnetic effects and so we need to consider the circuit itself to find the value of **B** at a point P near it. In this circuit, we introduce a network of wires as shown in Figure 7.21(b). Let each of these meshes carry the same

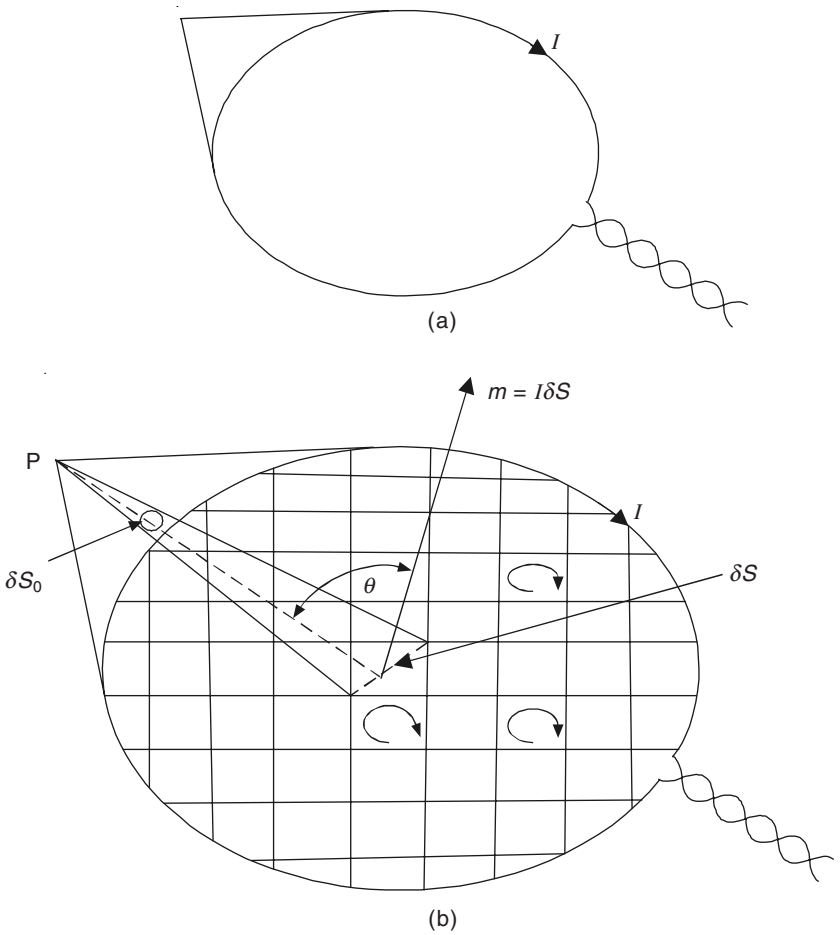


Figure 7.21 (a) The actual circuit and (b) the equivalent magnetic shell.

current I as in the original circuit and in the same sense; then each piece of wire would carry two equal and opposite currents, and hence there would be no resulting magnetic field due to it, except the outer wire of the original circuit which would carry the current I and thus produce a field at the point P.

If each mesh is fine enough, and has a small area δS whose linear dimensions are small

enough compared with the distance from the point P, then the net acts as a collection of current loop dipoles, each of moment (= $I\delta S = m$), and is called an 'equivalent magnetic shell'. This is because the net would produce the same magnetic field as a thin sheet of material magnetized with N poles on one side and S poles on the other, whose magnetic moment per unit area would be I.

$\therefore$ The magnetic potential $\delta\Omega$ due to a typical mesh is given by

$$\delta\Omega_m = \frac{m\cos\theta}{4\pi r^2} = \frac{I\delta S\cos\theta}{4\pi r^2} = \frac{I\delta S_0}{4\pi} \tag{7.51}$$

[See Eq. (1.54) for analogy with the electric dipole.]

where δS_0 is the solid angle subtended by the mesh δS at the point P.

$\therefore$ Due to the whole circuit

$$\Omega_m = \frac{IS_0}{4\pi} \tag{7.52}$$

where S_0 is the total solid angle subtended at P by the whole circuit or the shell.

7.9 DISTINCTION BETWEEN B AND H

In electrostatics, the primary vector is the electric force vector **E** defined in terms of the mechanical force on the unit charge. Subsequently when the Gauss' theorem is established, the auxiliary vector, i.e. the electric flux density vector **D** is introduced to express the idea of the flux emanating from each unit of electric charge. Similarly in the study of the magnetic fields, the primary vector is the magnetic flux density vector **B** defined in terms of the force due to the moving charge or the equivalent current loop. In the next step, when the magnetic circuit law is established, we introduce the secondary vector **H** which is associated with the source current through an imaginary contour. Thus **H** is connected with the currents which are the cause for the field; and **B** is associated with the force which is the effect of the magnetic field. So the distinction between these two vectors in the magnetostatics is that **H** can be considered as the cause and **B** as the effect.

In this regard, the names of the magnetic field vectors might appear somewhat confusing. The names for the electric field vectors seem quite logical; that is, **E** is the electric force due to the unit charge and **D** is the flux density vector associated with the flux emanating from the charges. But when it comes to the magnetic field vectors, we call **H** the magnetic intensity vector when **B** is the vector associated with the force. The reason for this apparent contradiction is that the development of the magnetostatics took place with the concept of the magnetic pole as the fundamental entity, in parallel with electrostatics. From this standpoint, **H** appears as the counterpart of **E**, and hence the two equations for these four vectors **E**, **D**, **B**, and **H** are written as

$$\mathbf{D} = \varepsilon_0\mathbf{E} \qquad \text{and} \qquad \mathbf{B} = \mu_0\mathbf{H} \tag{7.53}$$

in free space. The constants are somewhat analogous; μ_0 is called the 'absolute permeability of the free space', like ε_0 being the 'absolute permittivity of the free space'. Also just as we generalized ε_0 to $\varepsilon = \varepsilon_0\varepsilon_r$ for different dielectric media, we generalize for the permeability by $\mu = \mu_0\mu_r$ where μ_r is the relative permeability of the medium under consideration. We shall discuss these points in some detail when we consider the magnetic fields in media other than vacuum.

7.10 CALCULATION OF MAGNETIC FIELDS BY MEANS OF POTENTIAL AND THE MAGNETIC CIRCUIT LAW (AMPERE'S LAW)

In Sections 7.4 and onwards, we used the Biot–Savart's law to evaluate the magnetic fields due to a number of conductor configurations. Now we shall use the Ampere's law to find the magnetic fields due to some more configurations.

7.10.1 The Magnetic Field of Current in a Straight Circular Cylindrical Conductor

A straight circular cylindrical conductor of radius a carries a current I (Figure 7.22). We choose two concentric contours C_1 and C_2 as shown in Figure 7.22, to find **B** both outside and inside the conductor, respectively.

For the contour C_1:

$$\oint_{C_1} \mathbf{H} \cdot d\mathbf{l} = I$$

or

$$H 2\pi r = I$$

$$\therefore \qquad B_\theta = \mu_0 H = \frac{\mu_0 I}{2\pi r} \qquad r \geq a$$

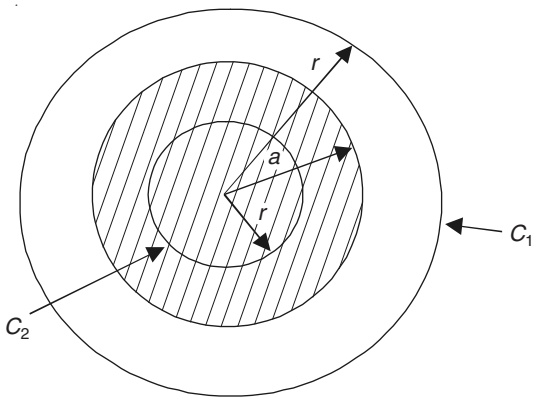


Figure 7.22 A straight circular conductor.

The direction of **B** is circumferential and that of the current is normal to the plane of the paper which contains the radii. Since, $J = \dfrac{I}{\pi a^2}$,

$$\mathbf{B} = \mu_0 \left(\frac{\mathbf{J} \times \mathbf{r}}{2} \right) \left(\frac{a}{r} \right)^2 \qquad r \geq a \tag{7.54}$$

For the field inside the conductor, we consider the contour C_2, and since C_2 encloses only a part of the current in the conductor,

$$\oint_{C_2} \mathbf{H} \cdot d\mathbf{l} = \left(\frac{I}{\pi a^2}\right)\left(\pi r^2\right) \qquad r \le a$$

$$\therefore \qquad B_\theta = \mu_0 H = \mu_0 \left(\frac{I}{2\pi r}\right)\left(\frac{\pi r^2}{\pi a^2}\right) = \frac{\mu_0 I r}{2\pi a^2} \qquad r \le a$$

$\therefore$ Expressing as vectors,

$$\mathbf{B} = \mu_0 \left(\frac{\mathbf{J} \times \mathbf{r}}{2}\right) \qquad r \le a \tag{7.55}$$

If now B_θ is plotted as a function of the radius (Figure 7.23), then it will be seen from Eqs. (7.55) and (7.54) that inside the conductor B_θ increases linearly from 0 at the centre ($r = 0$) to a

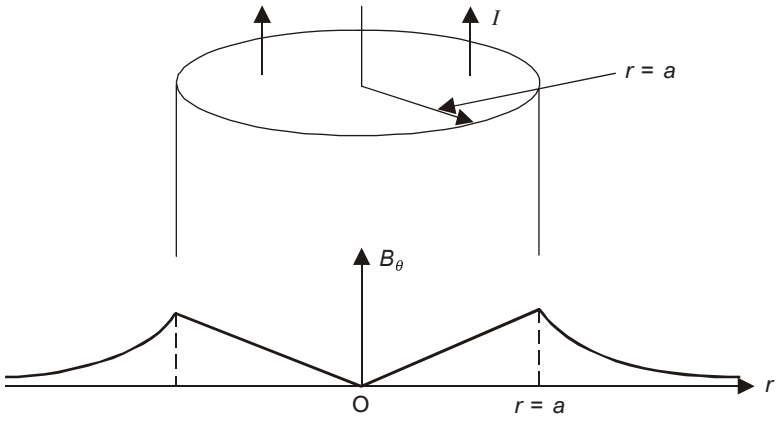


Figure 7.23 *B* due to a long straight current-carrying conductor.

maximum on the surface at $r = a$, and it starts decreasing from this maximum value inversely (or hyperbolically) as r increases, decreasing asymptotically to zero as $r \to \infty$.

7.10.2 The Magnetic Field of Current in a Coaxial Cable

The magnetic field of the current in a coaxial cable can be evaluated by exactly the same method, in all the three or four regions of the cross-section as shown in Figure 7.24.

The total current in the inner and the outer conductors are of the same magnitude, but opposite in directions.

By using the Ampere's law formula, i.e.

$$\oint_C \mathbf{H} \cdot d\mathbf{l} \; = \text{the enclosed current,}$$

we get the values of **B** in different regions as

$$B = \mu_0 \left(\frac{I}{2\pi a^2}\right) r \qquad \text{for } r \le a$$

$$B = \mu_0 \left(\frac{I}{2\pi r}\right) \qquad \text{for } a \le r \le b \tag{7.56}$$

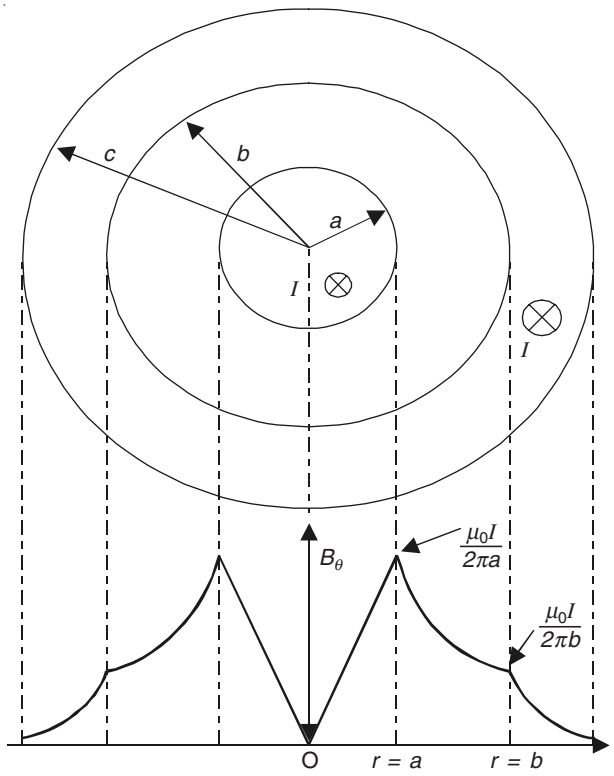


Figure 7.24 A coaxial cable.

$$B = \mu_0 \left(\frac{I}{2\pi r}\right)\left(\frac{c^2 - r^2}{c^2 - b^2}\right) \qquad \text{for } b \le r \le c$$

$$B = 0 \qquad \text{for } r \ge c$$

and in all the regions, the direction of **B** is circumferential. So expressing **B** in vectorial form,

$$\mathbf{B} = \mu_0 \left(\frac{\mathbf{J}_i \times \mathbf{r}}{2}\right) \qquad \text{for } r \le a$$

$$\mathbf{B} = \mu_0 \left(\frac{\mathbf{J}_i \times \mathbf{r}}{2}\right)\left(\frac{a}{r}\right)^2 \qquad \text{for } a \le r \le b \tag{7.57}$$

$$\mathbf{B} = \mu_0 \left(\frac{\mathbf{J}_i \times \mathbf{r}}{2}\right)\left(\frac{a}{r}\right)^2 + \mu_0 \left(\frac{\mathbf{J}_0 \times \mathbf{r}}{2}\right)\left(\frac{b}{r}\right)^2 - \mu_0 \left(\frac{\mathbf{J}_i \times \mathbf{r}}{2}\right) \qquad \text{for } b \le r \le c$$

where

$|\mathbf{J}_i| = \left(\frac{I}{\pi a^2}\right)$—the current density in the inner conductor

$|\mathbf{J}_o| = \left[\frac{I}{\pi(c^2 - b^2)}\right]$—the current density in the outer conductor.

As before, the only nonzero component of **B** is B_θ, and also the Gauss' theorem would confirm that $B_r = 0$ and $B_z = 0$. B_θ as a function of r has been plotted in the lower portion of Figure 7.24.

7.10.3 The Magnetic Field Inside a Cylindrical Circular Hole in a Cylindrical Circular Conductor

The conductor is carrying a current of uniform density **J**, the hole being drilled eccentrically with its axis parallel to the axis of the conductor (Figure 7.25). Though the problem sounds

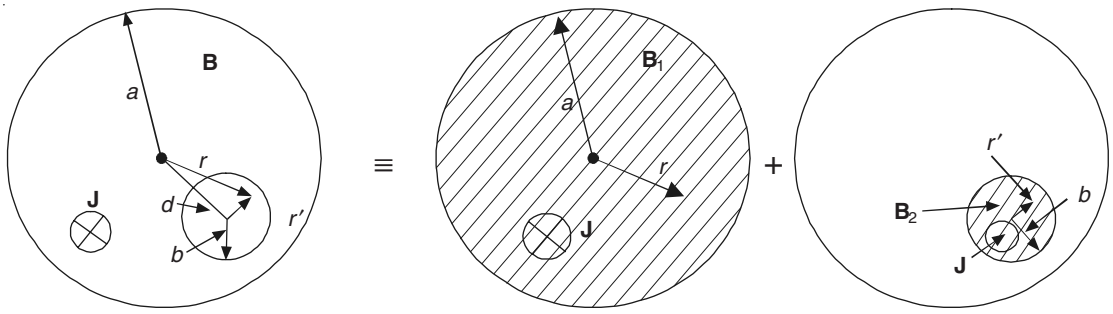


Figure 7.25 Circular conductor (with a hole) resolved into equivalent conductors.

rather complicated, basically it is a very simple problem. It can be solved directly by applying the Ampere's law and using the principle of superposition. We can use the result obtained in Eq. (7.55) of the problem in Section 7.10.1. The configuration of the hole shown in Figure 7.25 can be resolved into two equivalent circular conductors carrying the same current density but in opposite directions, as explained diagrammatically.

In each conductor, the magnetic flux density **B** in the cross-hatched region can be obtained by the Ampere's law, as in Eq. (7.55).

$$\mathbf{B}_1 = \mu_0\left(\frac{\mathbf{J}\times\mathbf{r}}{2}\right) \qquad \text{and} \qquad \mathbf{B}_2 = \mu_0\left(\frac{\mathbf{J}\times\mathbf{r}'}{2}\right)$$

$\therefore$ The resultant **B** in the region of the hole

$$\mathbf{B}_1 - \mathbf{B}_2 = \mu_0\left[\frac{\mathbf{J}\times(\mathbf{r}-\mathbf{r}')}{2}\right] = \mu_0\left(\frac{\mathbf{J}\times\mathbf{d}}{2}\right) \tag{7.58}$$

i.e. the magnetic field inside the cavity is *uniform.*

7.10.4 The Magnetic Field of Current in a Parallel Go-and-Return Circuit

Let us take two parallel thin wires, separated by a distance $2a$ and carrying currents $\pm I$. The geometry is as shown in Figure 7.26(a).

B at P due to go-current at A $= \mu_0\left(\frac{I}{2\pi \mathrm{AP}}\right)$, and its direction is at right angles to AP, towards the y-axis.

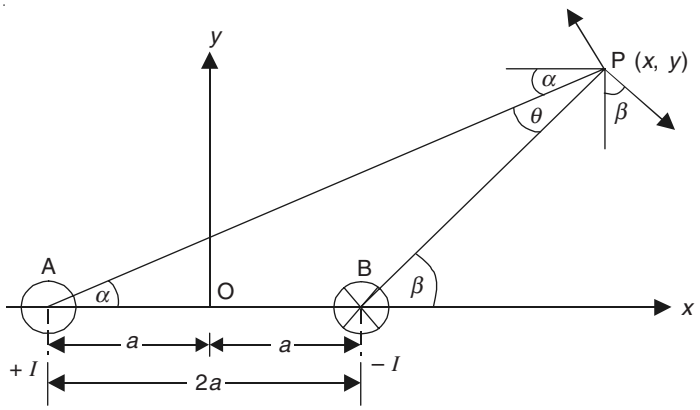


Figure 7.26(a) Magnetic field due to a parallel go-and-return circuit.

B at P due to return-current at B = $\mu_0\left(\frac{I}{2\pi \mathrm{BP}}\right)$, and its direction is at right angles to BP, towards the x-axis.

$\therefore$ The resultant B_x at P

$$= -\frac{\mu_0 I \cos\alpha}{2\pi\left[(x+a)^2 + y^2\right]^{1/2}} + \frac{\mu_0 I \cos\beta}{2\pi\left[(x-a)^2 + y^2\right]^{1/2}}$$

$$= -\frac{\mu_0 I (x+a)}{2\pi\left[(x+a)^2 + y^2\right]} + \frac{\mu_0 I (x-a)}{2\pi\left[(x-a)^2 + y^2\right]} \tag{7.59}$$

and the resultant B_y at P

$$= -\frac{\mu_0 I \sin\alpha}{2\pi\left[(x+a)^2 + y^2\right]^{1/2}} - \frac{\mu_0 I \sin\beta}{2\pi\left[(x-a)^2 + y^2\right]^{1/2}}$$

$$= -\frac{\mu_0 Iy}{2\pi\left[(x+a)^2 + y^2\right]} - \frac{\mu_0 Iy}{2\pi\left[(x-a)^2 + y^2\right]} \tag{7.60}$$

and $B_z = 0$.

The scalar magnetic potential at P is obtained by multiplying the current by the fraction of the view from P which is visible through the circuit, i.e.

$$\Omega_{\mathrm{P}} = \frac{I\theta}{2\pi} \tag{7.61}$$

where θ is the angle between AP and BP.

So to draw the equipotential through the point P, we have to maintain θ = constant, so as to obtain the equipotential. But θ remains constant when P moves round a circle through A and B (i.e. the angles in the same segment of a circle are equal). Hence the equipotentials will be a

system of circles passing through A and B, and they will have their centres on the *y*-axis, i.e. a 'coaxial system of circles' with their centres on the *y*-axis and all passing through the points A and B.

The lines of force everywhere would be orthogonal to the equipotentials, and hence will be another system of circles whose centres will lie on the line AB. It should be noted that these two systems of circles for this problem as shown in Figure 7.26(b) are the same as those obtained for

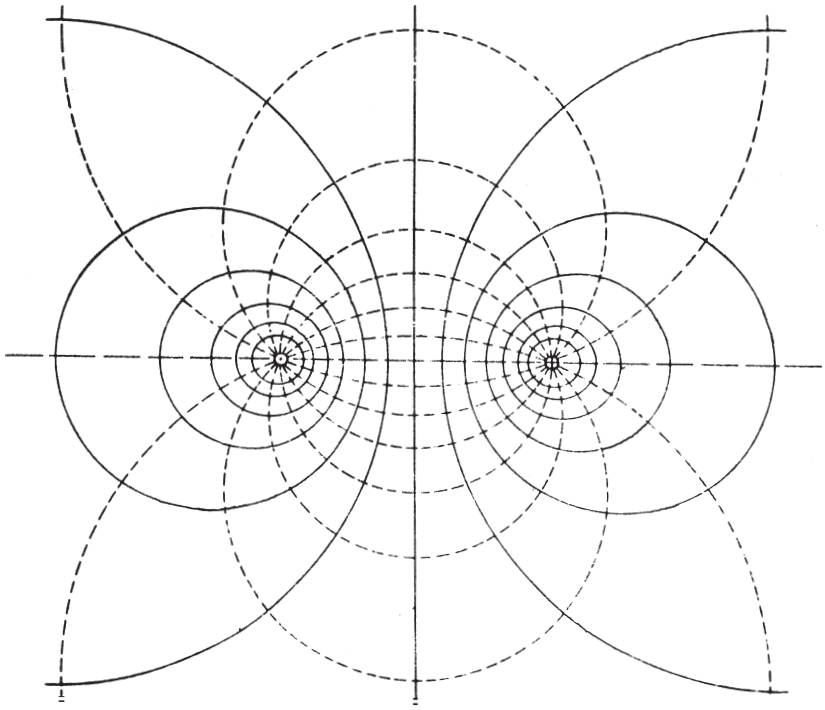


Figure 7.26(b) Flux plot about two equal and opposite parallel currents.

the field of the two parallel line charges, except that the lines of force and the equipotentials have got interchanged.

7.10.5 The Magnetic Field of a Toroidal Solenoid

Let us consider a toroid of rectangular cross-section (in fact, it can be of any cross-section), and a uniformly distributed winding on it of N turns; and let the current in the conductor (of the toroid) be I. For the contour C, shown in Figure 7.27,

$$\oint_C \mathbf{H} \cdot d\mathbf{l} = NI$$

or

$$H2\pi r = NI$$

$$\therefore \qquad H = \frac{NI}{2\pi r} \tag{7.62}$$

where r is the radius of the contour.

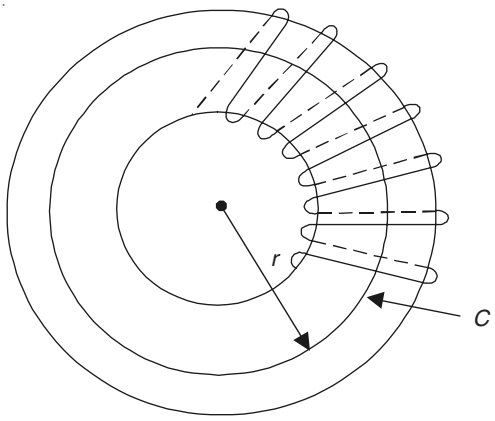


Figure 7.27 Toroidal solenoid.

For any other contour not lying within the volume of the ring, such a contour is not linked with the winding, and hence

$$\oint_C \mathbf{H}' \cdot d\mathbf{l} = 0 \qquad \text{or} \qquad \mathbf{H}' = 0$$

7.10.6 The Magnetic Field of an Infinitely Long Solenoid

This may be considered as the limiting case of a toroidal coil as discussed in Section 7.10.5.

We obtained for a toroid,

$$\mathbf{H} = \frac{NI}{2\pi r} = nI$$

where $n = \dfrac{N}{2\pi r}$ = number of turns/metre

$\therefore$ As r (= the radius of the contour) $\to \infty$, then

$$\mathbf{H} = nI \tag{7.63}$$

where n = turns/metre for an infinite solenoid.

Note: In Section 7.4.4, we obtained the magnetic field for a solenoid of finite axial length (by using the Biot–Savart's law), i.e.

$$H = \left(\frac{nI}{2}\right)(\cos\,\beta - \cos\,\alpha)$$

where α and β are as shown in Figure 7.9.

Hence for the infinitely long solenoid, $\beta = 0$ and $\alpha = \pi$.

$$\therefore \qquad H = nI$$

which is the same as Eq. (7.63).

It should be further noted that this is the magnetic field not only on the axis of the solenoid but right over the cross-section of the solenoid and is true for all cross-sections.

7.10.7 The Magnetic Flux Density on the Axis of a Circular Coil

We have already analyzed this problem by the Biot–Savart's law in Section 7.4.3. However, we shall now solve it by using the magnetic scalar potential. The circular coil is of radius a, carrying a current I, and we wish to find **B** at a point P on the axis of the coil, distant x from the centre (Figure 7.28).

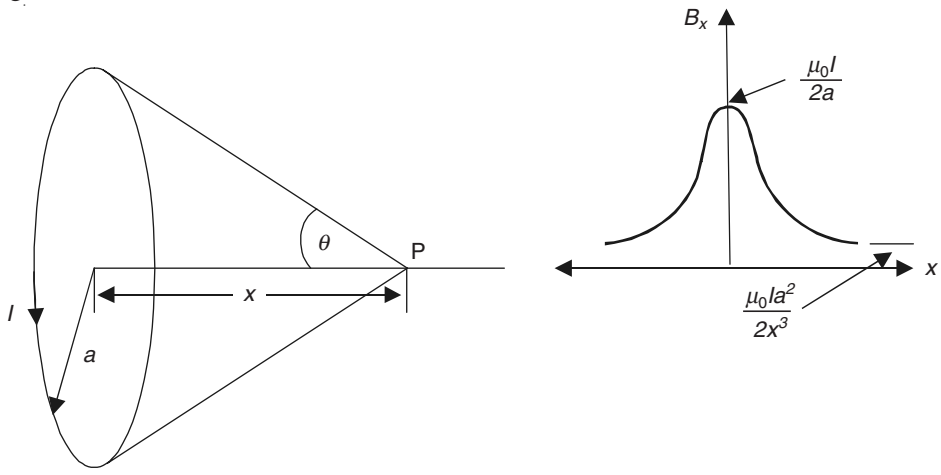


Figure 7.28 Magnetic field on the axis of a circular coil.

The scalar potential at the point P due to the current I in the circular coil is

$$\Omega = \frac{IS_0}{4\pi} \tag{7.64}$$

where S_0 is the solid angle subtended at the point P by the coil, and in this case, $S_0 = 2\pi(1 - \cos\theta)$.

$$\therefore \qquad \Omega = \frac{I(1-\cos\theta)}{2} = \frac{I}{2}\left(1 - \frac{x}{(x^2 + a^2)^{1/2}}\right) \tag{7.65}$$

$$\therefore \qquad B_x = -\mu_0 \frac{\partial \Omega}{\partial x} = \frac{\mu_0 I a^2}{2(x^2 + a^2)^{3/2}}, \qquad B_y = 0, \qquad B_z = 0 \tag{7.66}$$

At the centre of the coil, $x = a$,

$$\therefore \qquad B_x = \frac{\mu_0 I}{2a}$$

and as $x \to \infty$,

$$B_x = \frac{\mu_0 I a^2}{2x^3} \quad \text{or} \quad \frac{\mu_0 I A}{2\pi x^3}$$

where A is the area of the coil. The plot of B_x as a function of x is also shown in Figure 7.28.

7.10.8 Helmholtz Coils

If two circular coils of the same radius, and carrying identical currents in the same sense are

arranged coaxially, then their flux densities along the axis add up. A highly uniform magnetic field at the axial point mid-way between the coils, would be produced if the reduction in one field as we move away from this point is compensated by the increase in the other. This condition is achieved when the variation of B_x with x at the mid-point is linear, i.e. where $d^2B_x/dx^2 = 0$. When this is applied to Eq. (7.66), we get $x = a/2$. So the two coils have to be placed at a distance from each other, which is equal to the radius of the either coil. Such an arrangement is known as 'Helmholtz coils' (Figure 7.29).

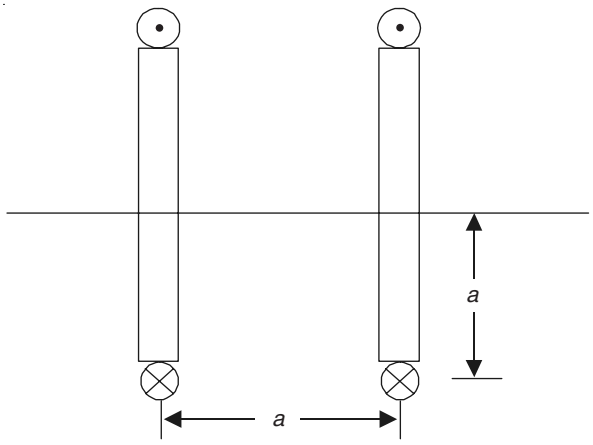


Figure 7.29 Helmholtz coils.

7.10.9 The Magnetic Field of a Planar Current Sheet

Let us assume that the surface current density over a very thin current sheet is J_S. By symmetry, the lines of **B** vector would be parallel to the sheet and normal to $\mathbf{J}_S$.

We apply the Ampere's law to the contour (abcda) shown in Figure 7.30(a), which is normal to the plane of the current sheet and also normal to the direction of the current.

$$\oint_{Cabcda} \mathbf{H} \cdot d\mathbf{l} = 2HL = J_S L$$

$$\therefore \qquad B = \frac{\mu_0 J_S}{2} \tag{7.67}$$

for the infinite planar current sheet.

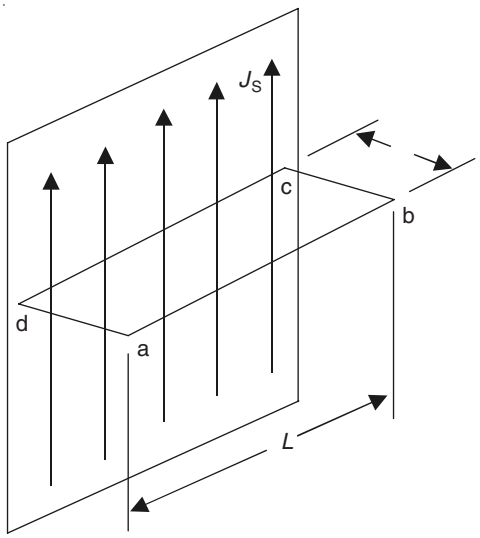
This value of magnetic field is independent of the distance from the current sheet. When there are two parallel current sheets of equal densities but of opposite directions, then by using the principle of superposition, the magnetic field between the sheets is

$$B = \mu_0 J_S$$

and that outside is

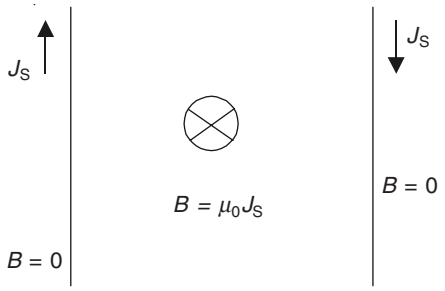
$$B = 0 \tag{7.68}$$

[See Figure 7.30(b)].



(a)

Figure 7.30(a) Planar current sheet.



(b)

Figure 7.30(b) Two parallel planar current sheets.

PROBLEMS

7.1 A circuit has the form of a regular hexagon in which the distance between the opposite vertices is $2a$. Prove that when the current in the circuit is I, the magnetizing force at its centre is $\left(\frac{\sqrt{3}}{\pi}\right)\frac{I}{a}$. (Specify the direction of this force).

7.2 Show that the magnetic induction at the centre of a loop of wire carrying a current I and shaped like a regular plane polygon of $2n$ sides, the distance between the parallel sides being $2n$, is $[\mu_0 nI/(\pi a)]$ sin $(\pi/2n)$.

7.3 Two similar concentrated circular coils are arranged on the same axis with their fields reinforcing each other. It is required that the magnetic field mid-way between them shall be as uniform as possible. Prove that the distance between the coils shall be equal to their radius.

7.4 In Problem 7.3, if the circular coils are replaced by square coils of side a, find the condition for similar uniformity of the field at the mid-point on the common axis of the coils.

7.5 A coil of negligible dimensions of N turns has the shape of a regular polygon of n sides inscribed in a circle of radius R metres. Show that the magnitude of the flux density at the centre of the coil, when it carries a current I, is $[\mu_0 NnI/(2\pi R)]\tan(\pi/n)$.

7.6 A magnetic field is so directed in a cylindrical coordinate system that the magnetic field intensity in a particular region is

(a) $\mathbf{H} = \mathbf{i}_\theta \dfrac{I}{2\pi r}$ (r is variable)

Show that in this region the electric current density $\mathbf{J}$ is zero.

(b) $\mathbf{H} = \mathbf{i}_\theta \dfrac{Ir}{2\pi R^2}$ (r is variable, R = constant)

Show that in the region where $\mathbf{H}$ is so expressed, there is a uniform current density. Find its magnitude.

8 Magnetic Field of Steady Currents in Presence of Magnetic Materials

8.1 BEHAVIOUR OF MAGNETIC SUBSTANCES IN THE MAGNETIC FIELD

8.1.1 Introduction

Atoms of all material bodies consist of a central heavy positive nucleus and a number of negative electrons revolving round it. This is a very crude, but from our point of view, an adequate model of the structure of matter. From the macroscopic point of view, these electrons, which are moving in their orbits with very high velocities, can be considered as tiny, atomic scale current loops. In addition, these electrons are also rotating about their axes which is known as the electron spin. So, if the electrons are considered as small charged balls, their spin, being a motion of the electric charge, can again be taken as equivalent to tiny closed current.

So if we are to analyze the effects of the substances on the macroscopic magnetic field into which they are introduced, then it is essential to analyze first the behaviour of a single current loop in a static magnetic field. Since these loops are quite small, it is quite justifiable to assume that they are in a uniform magnetic field. The most important effect is the torque by which a uniform magnetic field acts on a current loop. The torque tends to turn the loop so that the magnetic field of the loop increases the external magnetic field, with the result that the resultant magnetic field is greater than the primary magnetic field.

8.1.2 Torque on a Current Loop in a Uniform Magnetic Field

Since the atomic currents can be considered as equivalent to the elemental circular current loops, we will look at such a current loop of radius a, and carrying a current I, placed in a uniform magnetic field of flux density $\mathbf{B}$. Each moving charge inside the path is acted on by a force which is given by

$$\mathbf{F} = Q\mathbf{v} \times \mathbf{B}$$

Let the number of charges per unit volume be N and their average velocity be $\mathbf{v}$, then the magnetic force on the charges in an element of volume δv is

$$\delta \mathbf{F} = NQ\mathbf{v} \times \mathbf{B}\delta v \tag{8.1}$$

If this conducting region is a conductor of cross-sectional area δS and length δl, then

$$NQ\mathbf{v}\delta v = NQ\mathbf{v}\delta l\delta S = (NQv\delta S)\delta \mathbf{l} = (J\delta S)\delta \mathbf{l} = I\delta \mathbf{l} \tag{8.2}$$

where J is the magnitude of the current density, and $\delta \mathbf{l}$ is assumed to be in the direction of the current flow.

$\therefore$ Magnetic force on a small current element $I\delta \mathbf{l}$ is

$$= \delta \mathbf{F} = I\delta \mathbf{l} \times \mathbf{B} \tag{8.3}$$

$\therefore$ The total force on the contour is

$$\mathbf{F} = \oint_C I\delta \mathbf{l} \times \mathbf{B} = I \oint_C \delta \mathbf{l} \times \mathbf{B} = 0 \tag{8.4}$$

for any closed contour in a uniform magnetic field.

So Eq. (8.4) implies that such a force can cause no translation of the contour. But the elemental forces on the elements of the contour generally tend to rotate it. This can be seen as follows. Let us resolve the vector $\mathbf{B}$ into two components, i.e. one component $\mathbf{B}_n$ normal to the surface of the loop (Figure 8.1) and another orthogonal component $\mathbf{B}_P$ parallel to the plane of the contour (Figure 8.2). From Figure 8.1, it can be seen that $\mathbf{B}_n$ does not have any translational effect,

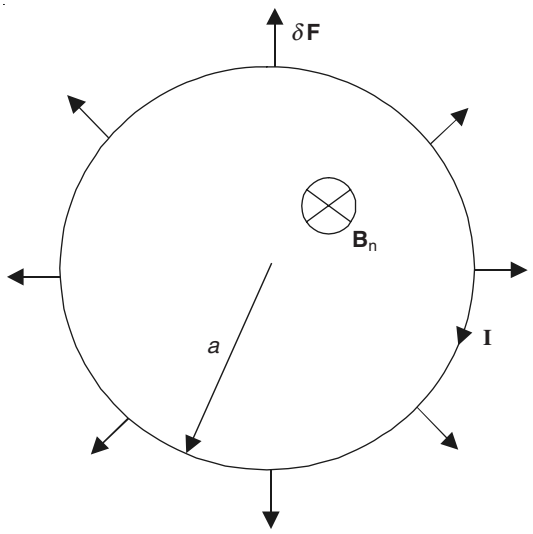


Figure 8.1 The total force on a closed current loop in a uniform magnetic field is zero.

but it tends to deform the contour (i.e. to increase its radius and hence a bursting effect). From Figure 8.2, it can be seen that the effect of the component $\mathbf{B}_P$ is to produce a rotating effect which tends to rotate the coil about the diametral axis OO′ which is perpendicular to the direction of $\mathbf{B}_P$. We shall now calculate this torque on the coil. So we consider two elements $\delta \mathbf{l}$ and $\delta \mathbf{l}'$ as shown in Figure 8.2, and calculate the elemental torque δT due to the elemental forces $\delta \mathbf{F}$ and $\delta \mathbf{F}'$ on $\delta \mathbf{l}$ and $\delta \mathbf{l}'$ respectively. Thus,

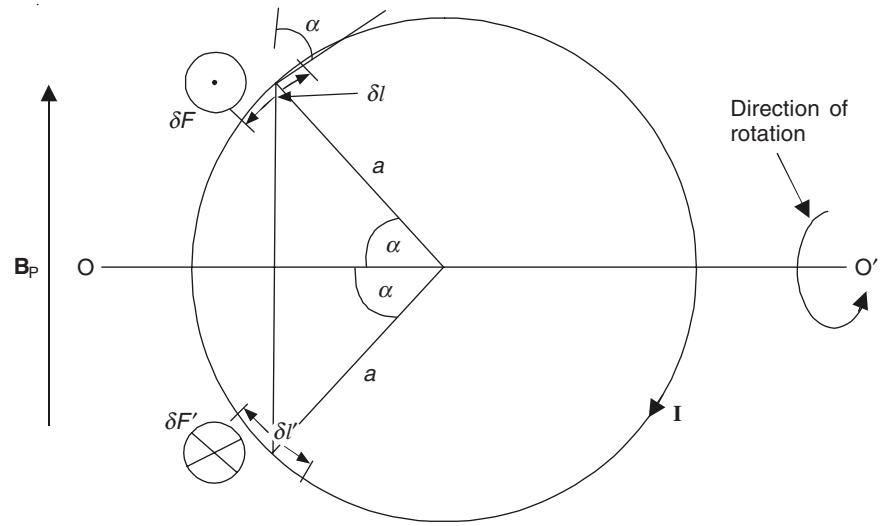


Figure 8.2 Torque of the magnetic forces on the closed current loop.

$$\delta T = \delta F a \sin\alpha + \delta F' a \sin\alpha$$
$$= 2\delta F a \sin\alpha = 2(I\delta l \sin\alpha B_P) a \sin\alpha = 2IaB_P \sin^2\alpha\, \delta l$$
$$= 2Ia^2 B_P \sin^2\alpha\, \delta\alpha$$

where $\delta l = a\delta\alpha$, and $\delta F = \delta F' = I\delta l B_P \sin\alpha$

$\therefore$ The total torque T is given by

$$T = \int_0^{\pi} 2Ia^2 B_P \sin^2\alpha\, d\alpha = 2Ia^2 B_P \frac{\pi}{2} = a^2 \pi I B_P \tag{8.5}$$

This can be expressed in terms of the total **B**, as is obvious from Figure 8.3, i.e. $\mathbf{B}_P = \mathbf{B} \sin\theta$, where θ is the angle between **B** and the positive normal to the plane of the loop.

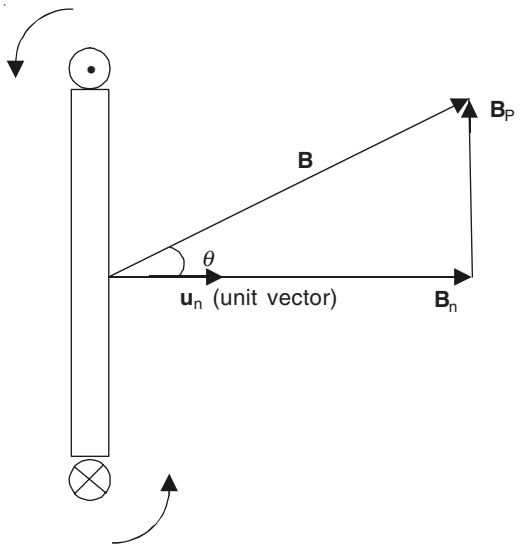


Figure 8.3 Torque of the magnetic force tending to align the vectors $\mathbf{u}_n$ and **B**.

$\therefore$ Representing the torque as a vector,

$$\mathbf{T} = a^2 \pi I \mathbf{u}_n \times \mathbf{B} = I\mathbf{S} \times \mathbf{B} \tag{8.6}$$

where $\mathbf{S} = \mathbf{u}_n \pi a^2$, the vector area covered by the coil.

The direction of the torque vector is determined according to the right-hand screw rule with respect to the contour rotation.

$$I\mathbf{S} = \text{magnetic moment of the loop} = \mathbf{m} \tag{8.7}$$

$$\therefore \qquad \mathbf{T} = \mathbf{m} \times \mathbf{B} \tag{8.8}$$

T tends to align the vectors **m** and **B**, since it (=**T**) is zero, when **m** and **B** are in the same direction, i.e. **T** tends to rotate the contour so as to make the flux of the external **B** through the contour maximal. The magnetic field of the contour then adds to the external magnetic field, thereby increasing the total flux density.

8.1.3 Behaviour of Magnetic Materials in a Magnetic Field

In a magnetic material by itself, the magnetic moments of the atoms are distributed in space statistically. Though a microscopic magnetic field exists in the vicinity of each atom, the net macroscopic field is zero. When the external magnetic field is applied, then each elemental loop which represents an atom, is acted upon by a torque, given by Eq. (8.8), i.e. $\mathbf{T} = \mathbf{m} \times \mathbf{B}$. This torque tends to align the moments (i.e. the loops) with the external magnetic field vector **B**. These loops, even when they are partially oriented, create a secondary macroscopic field. The result is that each loop (or the atomic current) is now situated in the external field and in this secondary macroscopic field of all other elemental currents. Thus the equilibrium state is a function of both the geometry of the external field, and the geometry and the physical properties of the magnetic material under consideration.

To determine the secondary macroscopic field, the distribution of the magnetic moments of the elemental current loops has to be found out. The secondary field is thus the field due to a given aggregate of the elemental circular currents in a vacuum. So the problem has thus been reduced to the evaluation of the flux density vector created by a single circular current loop and also **B** due to an aggregate of such loops.

8.2 FIELD OF AN ELEMENTARY CURRENT LOOP AND OF THE AGGREGATE OF THE LOOPS

We use the concept of the magnetic vector potential **A** to evaluate the magnetic field. We shall explain the concept of the potential later, and so these two sections can be left for later study though the results obtained here, can be used immediately. At this stage, we merely state that the vector potential **A** is related to the magnetic flux density **B**, by the relationship

$$\mathbf{B} = \text{curl } \mathbf{A} = \nabla \times \mathbf{A} \tag{8.9}$$

and that the vector potential due to the filamentary currents is

$$\mathbf{A} = \left(\frac{\mu_0}{4\pi}\right) \oint_C \left(\frac{I d\mathbf{l}}{r}\right) \tag{8.10}$$

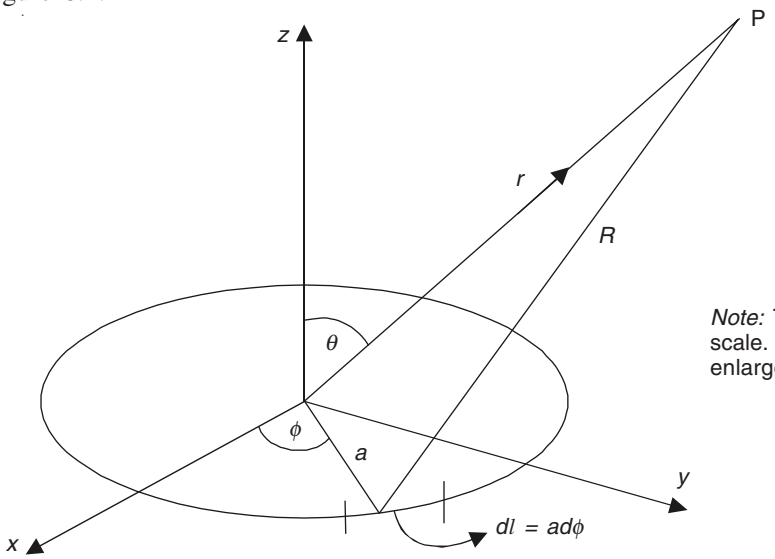
8.2.1 Field due to a Single Current Loop (Magnetic Dipole)

We consider a circular loop of radius a carrying a current I (see Figure 8.4), and we have to find the magnetic vector potential at a point P whose distance from the centre of the loop is r where $r >> a$. From Eq. (8.10), the lines of **A** are circles with the centres on the loop axis; and so **A** has only the ϕ-component (in the spherical polar coordinate system).

$\therefore$ The component dA_ϕ due to the element dl is

$$dA_\phi = \left(\frac{\mu_0 I}{4\pi}\right)\left(\frac{dl \sin \phi}{R}\right)$$

as shown in Figure 8.4.



Note: The figure is not to scale. The loop size is enlarged for clarity.

Figure 8.4 Vector potential due to a circular current element.

The point P is assumed to lie in the y-z plane ($x = 0$).

Now, $dl = ad\phi$, and

$$\frac{1}{R} \simeq \frac{1}{r - a \sin \phi \sin \theta} \simeq \frac{1}{r}\left(1 + \frac{a}{r} \sin \phi \sin \theta\right)$$

$$A_\phi = \left(\frac{\mu_0 Ia}{4\pi r}\right)\int_0^{2\pi}\left(1 + \frac{a}{r} \sin \phi \sin \theta\right)\sin \phi \, d\phi$$

$$= \frac{\mu_0 Ia^2 \pi \sin \theta}{4\pi r^2}$$

Using the definition of the magnetic moment **m** from Eq. (8.7),

$$\mathbf{A} = \frac{\mu_0 \mathbf{m} \times \mathbf{u}}{4\pi r^2} \tag{8.11}$$

and the magnetic flux density **B** is

$$\mathbf{B} = \text{curl } \mathbf{A} = \left(\frac{\mu_0}{4\pi}\right)\nabla \times \left(\frac{\mathbf{m} \times \mathbf{u}}{r^2}\right) \tag{8.12}$$

where

m = the magnetic moment of the loop

u = the unit vector in the direction of **r**.

8.2.2 Field of Aggregates of Loops (Magnetic Moment Density Vector or Magnetization Vector)

In a piece of magnetized material, the magnetic moments of all the elemental loops are known (say). Then the resulting vector potential at any point is

$$\mathbf{A} = \left(\frac{\mu_0}{4\pi}\right)\sum_{v} \left(\frac{\mathbf{m} \times \mathbf{u}}{r^2}\right) \tag{8.13}$$

where v is the volume of the magnetized piece. In the macroscopic theory, it is not necessary to know the magnetic moments of the individual loops. If the vector sum of the elemental magnetic moments inside any 'physically small' volume δv is determined, then the density of the magnetic moments is given by

$$\mathbf{M} = \frac{\text{sum of all the } \mathbf{m}\text{'s inside } \delta v}{\delta v} \tag{8.14}$$

This is the definition of the 'magnetization vector'. If the vector **M** is known at all the points, then the vector potential created by the small loops inside a volume v can be determined as an integral instead of the summation of Eq. (8.14). So a small volume δv has a moment

$$\sum_{\delta v} \mathbf{m} = \mathbf{M}\, dv$$

and the equation for the vector potential [i.e. (8.14)] becomes

$$\mathbf{A} = \left(\frac{\mu_0}{4\pi}\right)\iiint_{v} \frac{\mathbf{m} \times \mathbf{u}}{r^2}\, dv \tag{8.15}$$

So if **M** is evaluated as a function of the coordinates, then **A** and hence **B** can be determined easily. But the magnetization vector is a function of both the external magnetic field as well as the state of the induced magnetism at all the points, since this produces an additional magnetic field. Thus we are faced with a circular argument, i.e. the total magnetic field at a point inside a magnetized matter is a function of the magnetization throughout the volume under consideration, and the magnetization at a point is a function of the total magnetic field. So the analysis of the transient process becomes extremely complicated. But under the final equilibrium state which is to be determined, the dependence of **M** on the total **B** must be known. For most substances, the magnetization at a given point is a linear function of the total **B** at that point, expressed in the form

$$\mathbf{M} = \left(\frac{1}{\mu_0}\right)\left(\frac{\chi_m}{1+\chi_m}\right)\mathbf{B} \tag{8.16}$$

χ_m is a dimensionless constant, known as the magnetic susceptibility of the material considered.

For all linear magnetic materials $\chi_m << 1$. But for the ferromagnetic materials, $\chi_m >> 1$ and is a function of both **B** and the magnetic history of the specimen.

8.3 EQUIVALENCE OF THE MACROSCOPIC CURRENTS TO THE MAGNETIZED SUBSTANCES

We have thus seen that the magnetic field of a magnetized substance is due to the elemental current loops which represent the magnetic properties of the individual atoms. These currents can be considered as producing certain macroscopic currents. We shall now express the density of the macroscopic currents resulting from the microscopic currents in terms of the magnetization vector **M**.

We consider a part of the surface of a magnetized material. It can be easily seen that for uniform magnetization, the macroscopic resultant of the elemental currents inside the material would be zero. But for a thin layer near the surface, the uncompensated currents exist which would result in a thin current sheet of thickness same as the diameter of the current loops, and hence can be regarded as a 'surface current'. The density J_{mS} of the surface current can be determined as follows. Let there be N elemental loops per unit volume, and the radius of each atomic loop is a, carrying a current I. Then the surface current density is

$$J_{mS} = INa^2 \pi = Nm = M \tag{8.17}$$

i.e. it is equal to the magnetization vector $|\mathbf{M}|$.

If, however, these loops are inclined to the outer surface by an angle α (instead of being normal), then

$$J_{mS} = INa^2 \pi \sin \alpha = M \sin \alpha$$

So we can represent either of the above situations by

$$\mathbf{J}_{mS} = \mathbf{M} \times \mathbf{u}_n \tag{8.18}$$

where $\mathbf{u}_n$ is the unit vector normal to the magnetized surface, directed away from it.

When the magnetization of a magnetized body is uniform, the resultant currents appear on the surface only. If the magnetization is not uniform, the resultant macroscopic currents can appear throughout the body. By a similar line of argument, considering similar contours, we can show that in a rectangular contour

$$\mathbf{J}_m \cdot \mathbf{u}_n = \mathbf{u}_n \cdot \text{curl } \mathbf{M} \tag{8.19}$$

where $\mathbf{J}_m$ is the density of the macroscopic resultant of the microscopic currents.

$$\text{Or} \qquad \mathbf{J}_m = \text{curl } \mathbf{M} \tag{8.20}$$

So the total magnetic field can now be analyzed in terms of the field of the original macroscopic currents and the secondary equivalent currents, both existing in vacuum. We have then reduced the problem to the analysis of a system of currents situated in vacuum—a situation which has been discussed in Chapter 7.

8.4 THE GENERALIZED FORM OF AMPERE'S CIRCUITAL LAW

Thus according to the equivalence established in Section 8.3, the secondary sources of the

magnetic field inside the magnetized material can be reduced to a system of macroscopic currents. The real currents and these equivalent currents are then considered to exist in a vacuum. Suppose, now at a point, both the real currents (=**J**) and the macroscopic resultant of the microscopic currents exist. Hence the total current creating the magnetic field is the sum of these components, i.e.

$$\mathbf{J}_{\text{total}} = \mathbf{J} + \mathbf{J}_{\text{m}} = \mathbf{J} + \text{curl } \mathbf{M} \tag{8.21}$$

from Eq. (8.20).

Then the Ampere's circuital law has to be modified for $\mathbf{J}_{\text{total}}$ rather than for **J** only, i.e.

$$\oint_C \mathbf{B} \cdot d\mathbf{l} = \mu_0 \iint_S (\mathbf{J} + \nabla \times \mathbf{M}) \cdot d\mathbf{S} \tag{8.22}$$

Applying the Stoke's theorem to the second term on the right-hand side and rearranging

$$\oint_C \left(\frac{\mathbf{B}}{\mu_0} - \mathbf{M} \right) \cdot d\mathbf{l} = \iint_S \mathbf{J} \cdot d\mathbf{S} \tag{8.23}$$

The significance of the left-hand side term is as follows. The line integral of the difference of the two vectors is thus always equal to the 'real' current crossing a surface enclosed by the contour. This quantity is then defined as the vector **H** called the magnetic field intensity vector. (This is the generalized definition of **H** when any magnetized material is present.) That is,

$$\mathbf{H} = \frac{\mathbf{B}}{\mu_0} - \mathbf{M} \tag{8.24}$$

and so Eq. (8.23) becomes

$$\oint_C \mathbf{H} \cdot d\mathbf{l} = \iint_S \mathbf{J} \cdot d\mathbf{S} \tag{8.25a}$$

This is the generalized form of the Ampere's law. Sometimes it is more convenient to write it in a more explicit form as

$$\oint_C \mathbf{H} \cdot d\mathbf{l} = \text{sum of all the currents through } C \tag{8.25b}$$

The differential form of this law is

$$\text{curl } \mathbf{B} = \mu_0(\mathbf{J} + \text{curl } \mathbf{M})$$

or

$$\nabla \times \left(\frac{\mathbf{B}}{\mu_0} - \mathbf{M} \right) = \mathbf{J} \tag{8.26}$$

Using the definition of **H** from Eq. (8.24), this becomes

$$\nabla \times \mathbf{H} = \mathbf{J} \tag{8.27}$$

This generalized form of the Ampere's law as defined by Eq. (8.27) is valid for the magnetic fields of the steady currents only. In fact, this is a restricted form of another of the "Maxwell's electromagnetic field equations". Later we shall see how this equation has to be generalized further by including an extra term to account for the effects of the time-varying magnetic fields.

For linear magnetic materials, the magnetization vector **M** is proportional to the magnetic flux density vector **B** and hence by Eq. (8.24), it is also proportional to the magnetic field intensity vector **H**. This linear relationship is expressed as

$$\mathbf{M} = \chi_m \mathbf{H} \tag{8.28}$$

As already noted before in Section 8.2.2, χ_m is the magnetic susceptibility of the material. Combining Eq. (8.28) with Eq. (8.24), we get

$$\mathbf{B} = \mu_0 (\mathbf{H} + \mathbf{M}) = \mu_0 (1 + \chi_m)\mathbf{H} \tag{8.29}$$

The quantity $(1 + \chi_m)$ is denoted by μ_r and is known as the *relative permeability* of the medium. The product $\mu_0 \mu_r$ is denoted by μ and is called the *permeability* of the medium. Thus,

$$(1 + \chi_m) = \mu_r, \quad \mu_0 (1 + \chi_m) = \mu_0 \mu_r = \mu \tag{8.30}$$

and Eq. (8.29) becomes

$$\mathbf{B} = \mu \mathbf{H} \tag{8.31}$$

Thus the magnetization vector **M** inside a linear magnetic material can be expressed in terms of **B** or **H** and μ as

$$\mathbf{M} = \frac{\mathbf{B}}{\mu_0} - \mathbf{H} = \left(\frac{\mu - \mu_0}{\mu \mu_0} \right) \mathbf{B} = \left(\frac{\mu}{\mu_0} - 1 \right) \mathbf{H} \tag{8.32}$$

In Table 8.1 below are given the values of μ_r and χ_m for some linear substances as well as for some nonlinear highly magnetizable substances. Note that both μ_r and χ_m are dimensionless.

Table 8.1 μ_r and χ_m for some common substances

Substance	μ_r	χ_m	*Substance*	μ_r	χ_m
Bismuth	0.99983	-1.7×10^{-4}	Air	1.00000036	3.6×10^{-7}
Silver	0.99998	-2.6×10^{-5}	Aluminium	1.000021	2.1×10^{-5}
Copper	0.99999	-0.94×10^{-5}	Cobalt	250	nonlinear
Water	0.99999	-0.88×10^{-5}	Nickel	600	nonlinear
Vacuum	1.00000	0.00	Soft iron	5000	nonlinear

It should be noted that μ_r can be less than unity which was not the case with the relative permittivity. Substances with $\mu_r < 1$ are known as 'diamagnetic', and those with $\mu_r > 1$ but almost equal to unity are called 'paramagnetic'. Substances with $\mu_r >> 1$ which are always nonlinear are called 'ferromagnetic'. We shall now briefly discuss the effects of the externally applied magnetic field on these substances.

8.5 EFFECT OF AN EXTERNALLY APPLIED MAGNETIC FIELD ON MATERIAL SUBSTANCES

Before we consider this effect on the different classes of materials just described, we shall write down the expression for the 'magnetic dipole moment' (=**m**) in terms of the parameters of the circulating electrons in the atom. The dipole moment is defined by the product of the current and the area of the orbit as

$$|\mathbf{m}| = \left(\frac{ev}{2\pi R}\right)\pi R^2 = \frac{evR}{2} = \frac{e\omega R^2}{2} \tag{8.33}$$

where e is the charge of the electron orbiting with a velocity of magnitude v in an orbit of radius R. At a distance from the electron this movement appears as an average current of ($ev/2\pi R$) amperes. This quantity is called the *magnetic dipole moment* because it is a product of a current and an area, and is thus analogous to the electric dipole moment which is a product of the charge and the distance.

Expressing the magnetic dipole moment as a vector,

$$\mathbf{m} = -\left(\frac{e}{2}\right)(\mathbf{v} \times \mathbf{R}) \tag{8.34}$$

and so we have the magnetization vector (or the magnetization per unit volume) given before by Eq. (8.14) as

$$\mathbf{M} = N\mathbf{m}$$

where N is the number of such effective dipoles per unit volume.

Next we consider the effect of an externally applied magnetic field on the material substances.

Diamagnetic substances. Those substances which possess atomic structures so that no motion of the plane of the orbits occurs under the influence of an externally applied field are called *diamagnetic* substances. For such substances, an ordered array of the orbital planes produces no external magnetic effect caused by the substance atoms when no external field is applied. But when there is an external magnetic field, then those orbits so located that the normal to their areas have a component aligned with the applied field, would produce a reaction though there will be no reorientation of their planes. Let us consider an electron in an orbit whose plane is normal to the direction of the externally applied field $\mathbf{B}_0$. When the external field $\mathbf{B}_0$ is not present, the orbit is determined by the balance of the coulomb attractive force between the electron and the positively charged nucleus, and the centrifugal force acting outwards [Figure 8.5(a)]. Equating the magnitudes of these forces,

$$\frac{e^2}{4\pi\varepsilon_0 R_0^2} = m_e \omega_0^2 \, R_0 \tag{8.35}$$

where

e = electron charge, in coulombs

ε_0 = permittivity of free space, in henries/m

m_e = mass of the electron, in kg

ω_0 = original angular velocity of the electron, in rads/s.

From this equation, it is seen that $\omega_0^2 R_0^3$ is a constant if the mass of the electron is constant.

When an external field $\mathbf{B}_0$ is applied, as shown in Figure 8.5(b), under the effects of the additional field $\mathbf{B}_0$, and the velocity $\mathbf{v}$, the electron will experience a larger outwardly directed force (due to this magnetic field). This (centrifugal force) has to be balanced by a larger inwardly directed force, which is given by

$$\frac{e^2}{4\pi\varepsilon_0 R^2} = m_e \omega^2 R + e\mathbf{B}_0 (R\omega) \tag{8.36}$$

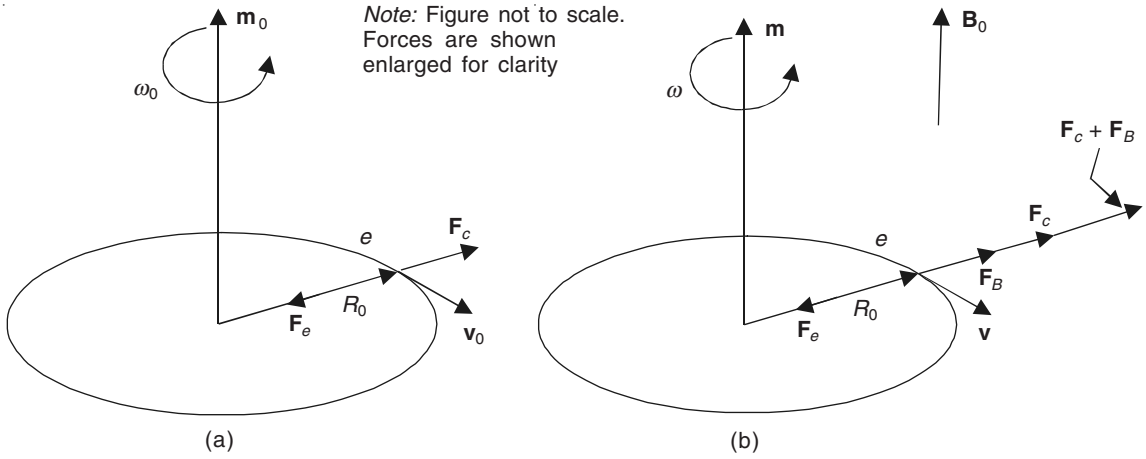


Figure 8.5 Orbiting electron in the absence and the presence of an external field $\mathbf{B}_0$.

where R and ω are the new values of the radius and the angular velocity, respectively. Calculating $\omega^2 R^3$ from the above equation and using Eq. (8.35), we get

$$\omega_0^2 R_0^3 - \omega^2 R^3 = \left(\frac{e}{m_e}\right) B_0 R^3 \omega \tag{8.37}$$

i.e. $\omega^2 R^3$ is less than $\omega_0^2 R_0^3$ as a result of the externally applied magnetic field, and so the magnetic moment **m** also would be smaller than $\mathbf{m}_0$ (i.e. the value of **m** when there is no externally applied magnetic field). The effect of $\mathbf{B}_0$ has been to produce a change in **m** in the direction opposite to that of $\mathbf{B}_0$. This is due to the change in **v** and assuming that the change in **m** is from the changes in ω and not through R_0, it can be shown that the change in **m** is given by

$$\Delta\mathbf{m} = -\left(\frac{e^2 R_0^2 \mu_0^2}{4m}\right)\mathbf{H} \tag{8.38}$$

and multiplying this equation by N and then comparing it with Eq. (8.28), we get

$$\chi_{\mathrm{m}} = -\left(\frac{e^2 R_0^2 \mu_0^2}{4m}\right) \tag{8.39}$$

This shows that the diamagnetic materials have negative susceptibilities. This is an approximate analysis, and for a more rigorous explanation, we have to go into the quantum mechanical considerations. However for the approximate values assigned to N and R_0, we find that χ_{m} for the diamagnetic substances lies between -10^{-6} and -10^{-4}. For approximate calculation purposes for the diamagnetic materials, μ_{r} may be taken as unity.

Paramagnetic substances. These substances exhibit small positive susceptibilities caused by the spin characteristics of the electrons. The magnetic moment **m** is a function of the temperature of the substance, i.e.

$$\chi_{\mathrm{m}} = +\frac{C}{T} \tag{8.40}$$

where C is a characteristic constant of the substance, called the Curie constant. This equation is

valid over a reasonable range of temperatures, but does not apply at very low temperatures. Substances, such as oxygen, various oxides and chlorides possess the property of paramagnetic action, and χ_m has a value in the range of $+10^{-3}$.

Ferromagnetic substances. In the previous two types of substances, i.e. the diamagnetic and the paramagnetic, there is negligible interaction between the individual atomic moments. But there is another class of substances, like iron, cobalt, nickel, and their alloys, in which there occurs a very strong interaction among the groups of the atomic moments. Regions in which this interaction occurs are called 'domains' and constitute groups of an order of 10^{15} atoms. The linear dimensions of a domain are of the order of 10^{-6} m or about 20 microns (=μ), and the domains are separated by the walls about 50 Å (Angstroms) thick in which the direction of the spins is changing. When there is no external applied magnetic field, each domain may have its spins aligned along one of the several possible axes, depending on the particular substance. At this stage, we differentiate between two types of ferromagnets: (a) soft, which retain no magnetism when there is no external magnetic field and (b) hard, which may form permanent magnets.

Qualitatively, the behaviour of these two types of ferromagnets can be explained as follows. (See Figure 8.6.) Because the domains can possess random orientations, the net residual magnetic flux density **B** in the material may be small, as in the so called soft substances like pure iron. Other substances, like certain alloys, possess easy and hard directions of magnetization. When an external magnetic field is applied in the direction of easy magnetization, the domains align gradually as the field strength is increased. The wall movement of these domains permit the growth of the size of these domains till the whole specimen becomes essentially a single domain. Under this condition, the material is said to be magnetically 'saturated'. When the magnetic field

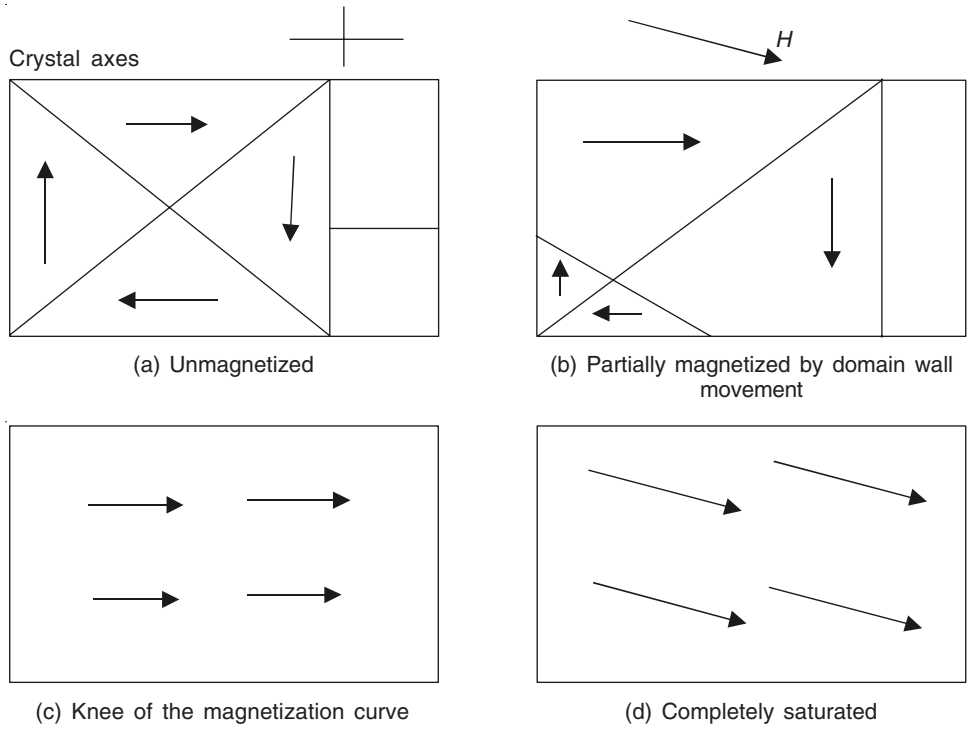


Figure 8.6 Domain movements in ferromagnetic substances during the application of an external magnetic field.

intensity is reduced, the domain walls may begin to move again, producing domains of smaller size and random orientations. Depending on the substance, a greater or lesser degree of magnetization may remain. If a considerable amount of magnetization remains, when the external magnetic field is reduced to zero, the substance is called a 'permanent magnet'. The details of such substances will be discussed later.

A study of the magnetizing and the demagnetizing action shows that this process is not a smooth one, and that the sudden motion of the domain walls produces an effect of slight discontinuity which is known as the 'Barkhausen effect'.

In ferromagnetic substances, there is a hysteresis effect occurring as the material experiences magnetization and demagnetization. Because of this and the nonlinear relationship between B and H for the material, a simple tabulation of χ or μ_r for ferromagnetic substances is not possible. The details of all such aspects will be explained later.

8.6 MAGNETIC FIELD INTENSITY VECTOR AND ITS INTERPRETATION

The lines of the magnetic field intensity vector **H** are defined as the geometric lines having the property that the vector **H** is tangential to them at all the points. But while representing, the magnetic field lines of **H** are at a disadvantage compared to the lines of the vector **B**, especially inside the magnetic materials. This is because **B** is divergenceless everywhere, but this is not so with **H**. In fact inside the materials

$$\text{Div } \mathbf{H} = -\text{ div } \mathbf{M}$$

$\therefore$ The vector **H** has sources and sinks, i.e. there are points at which some **H** lines originate and also the points at which these lines end. Also, according to Eq. (8.24), the magnetic field intensity vector **H** is a difference of two physically different quantities, i.e. $(\mathbf{B}/\mu_0)$ is proportional to the magnetic flux density vector at the point under consideration, due to both the macroscopic and the microscopic currents, and **M** is the density of the magnetic moments of the elemental atomic currents at that point. Thus **H** is a hybrid vector function made up of two components, having a completely different physical meaning. It is defined as a convenient quantity for analyzing the magnetic field in the presence of the magnetic materials.

As an example to illustrate this point, we study the following simple problem. We have the magnetic force component of the Lorentz force from Eq. (7.24) as

$$\mathbf{F} = Q\mathbf{v} \times \mathbf{B}$$

But since in vacuum, $\mathbf{B} = \mu_0\mathbf{H}$, we could also have written this component of the force as

$$\mathbf{F} = \mu_0 Q\mathbf{v} \times \mathbf{H}$$

If we consider a moving charge inside a permanent magnet, so that the interaction between the charge and the atoms of the magnet can be neglected, then from our standpoint, the charge would be deflected by the law of the force, $\mathbf{F} = Q\mathbf{v} \times \mathbf{B}$. But the other approach suggests that the deflection would be according to the law $\mu_0 Q\mathbf{v} \times \mathbf{H}$ in the opposite direction, since inside the magnet **B** and **H** are in opposite directions (as will also be seen later). Experimentally it is found that the force law is

$$\mathbf{F} = Q\mathbf{v} \times \mathbf{B}$$

8.7 BOUNDARY CONDITIONS (SURFACES OF DISCONTINUITY)

The boundary conditions are similar to those discussed for the electrostatics and the steady currents. The conditions are now the relations for the vectors **B** and **H** at the two adjacent points on the two sides of an interface between the two different magnetic surfaces (Figure 8.7).

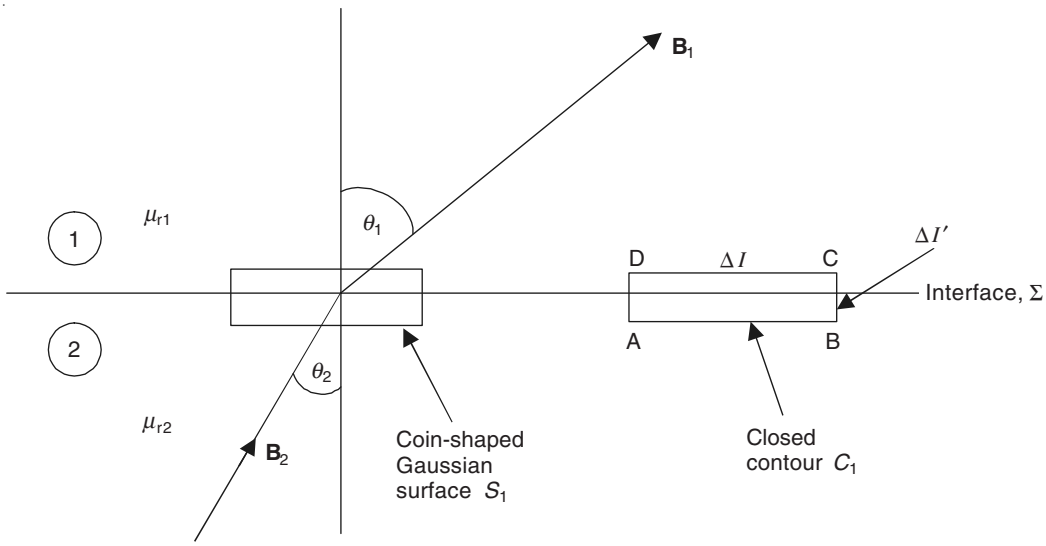


Figure 8.7 Interface between two different magnetic media.

We consider the interface, Σ, between the two different magnetic media (1) and (2) of relative permeabilities μ_{r1} and μ_{r2}, respectively. On this interface, we take a coin-shaped Gaussian surface of cross-sectional area δS_1 and find the outward flux of **B** from S_1. This flux must be zero because of the solenoidal property of **B**, i.e. div **B** = 0. (This is an alternative way of applying the 'principle of conservation of magnetic flux'.) So,

$$(B_1 \cos\ \theta_1 - B_2 \cos\ \theta_2)\delta S_1 = 0$$

as the flux out of the peripheral edge of the closed surface becomes zero in the limit.

$$\therefore \qquad B_1 \cos\ \theta_1 = B_2 \cos\ \theta_2 \qquad \text{or} \qquad B_{n1} = B_{n2} \tag{8.41}$$

i.e. the normal component of **B** is continuous across such a surface of discontinuity.

We next consider a closed contour C_1, i.e. ABCDA on the interface Σ. If there is no surface current on the interface, then

$$\oint_{\text{ABCDA}} \mathbf{H} \cdot d\mathbf{l} = 0$$

or

$$(H_1 \sin\ \theta_1 - H_2 \sin\ \theta_2)\Delta l + \text{contribution due to } \Delta l' = 0$$

The second term in the above equation $\rightarrow 0$ in the limit.

$$\therefore \qquad H_1 \sin\ \theta_1 = H_2 \sin\ \theta_2 \qquad \text{or} \qquad H_{t1} = H_{t2} \tag{8.42}$$

This can be written as

$$\frac{B_{t1}}{\mu_{r1}} = \frac{B_{t2}}{\mu_{r2}} \tag{8.43}$$

The above equation implies that if there is no surface current on the surface of discontinuity, then the tangential component of the vector **H** is continuous. Combining the above three equations, we get

$$\frac{\tan \theta_1}{\tan \theta_2} = \frac{\mu_{r1}}{\mu_{r2}} \tag{8.44}$$

This is the law of refraction of the magnetic field lines.

Let us consider an important practical case, when the medium (1) is air, so that $\mu_{r1} \simeq 1$ and $\mu_1 = \mu_0$, and the medium (2) is some ferromagnetic substance, such that $\mu_{r2} >> \mu_{r1}$. In that case,

$$\frac{\tan \theta_1}{\tan \theta_2} = \frac{\mu_{r1}}{\mu_{r2}} \simeq 0 \tag{8.45}$$

in which case $\theta_1 = 0$, i.e. the magnetic flux density vector **B** in air is practically normal to the surfaces of ferromagnetic materials.

Next, we consider again the interface of discontinuity, but now with a surface current I_S per unit width (Figure 8.8).

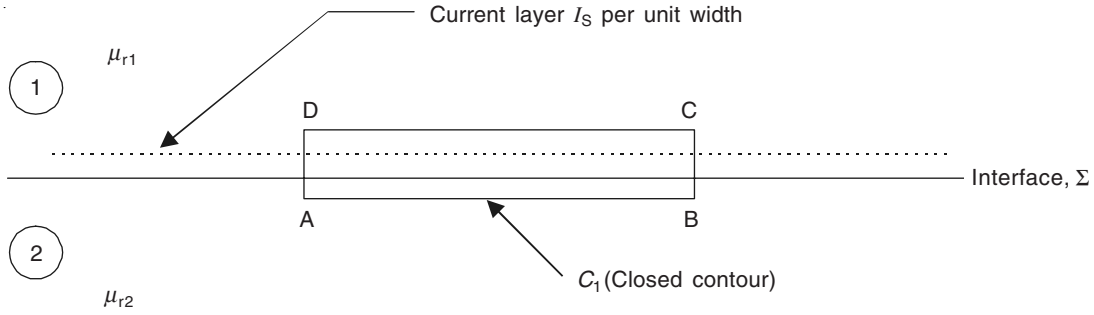


Figure 8.8 Surface of discontinuity with surface current on it.

In this case as well, the normal component of **B** would be continuous, for the same reason as before. But when we consider the closed contour ABCDA,

$$(H_{t1} - H_{t2})\,\Delta l = I_S\,\Delta l \qquad \text{or} \qquad H_{t1} - H_{t2} = I_S \tag{8.46}$$

i.e. the tangential component of **H** is discontinuous now. Furthermore if the magnitude of I_S is large enough, the direction of H_t might get reversed as one passes across the interface plane Σ.

8.8 THE MAGNETIC CHARACTERISTICS OF IRON (FERROMAGNETIC MATERIALS)

As we have seen earlier, the magnetic materials are divided into three main groups. The first two groups, i.e. the diamagnetic and the paramagnetic substances, have their relative permeabilities

very near unity (due to very small negative or positive values of their susceptibilities) and hence can be considered to be having practically no influence on the magnetic field. They are seldom used, except for special purposes, and so we shall not discuss them here. The third group of materials are known as ferromagnetic materials. They are iron, cobalt, nickel, and their chemical compounds and alloys. These have very large relative permeability, and are nonlinear and their magnetic properties are functions of their magnetic history. They are very widely used in electrical and electronic engineering, and so a basic knowledge of their important properties is essential. A complete analysis of their properties is highly complex, and for those readers who are interested, they will find the relevant information in specialist's books, such as *Ferromagnetism* by Bozorth. Here we shall describe the basic properties of these materials in brief.

So far we have been trying to build up a picture of magnetization of iron in terms of cause and effect. The cause is specified in terms of the magnetizing force **H**, and the effect in terms of **M** or **B** since these parameters are related by the equation

$$\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M}) = \mu_0(1 + \chi_m)\mathbf{H}$$

and for iron, this relationship is nonlinear. We will have a look at the **B**-**H** relationship for iron in two stages. First, we shall study the relationship by an example, and then the general relationship.

8.8.1 A Short Solid Iron Cylinder in a Straight Cylindrical Coil of Infinite Length

The axis of the iron cylinder is taken parallel to the axis of the coil. The cross-section of the iron cylinder is taken about one-tenth that of the coil. In this arrangement, the flux density **B**, far away from the iron, will be of uniform value, the same as for an air-cored solenoid, i.e. $B_0 = \mu_0 H = \mu_0 nI$ where n is the number of turns per unit axial length of the solenoid, and I is the current per turn. But when the flux approaches the iron, most of it tends to pass through the iron. The **B** pattern is shown in Figure 8.9(a). This can be considered to have been produced by superimposing the field of a magnetized bar [Figure 8.9(b)] on the uniform field due to the solenoid itself. Figure 8.10 shows the corresponding **H** lines. It will be seen that the **B** and the **H** lines in the air-cored solenoid are very similar, because **B** and **H** in the air are proportional. But in the iron part, the density of the **H** lines is extremely low [Figure 8.10(a)]. This is because if we study the **H** lines in the magnetized bar [Figure 8.10(b)], it will be seen that the **H** lines in the bar are in the opposite direction. So considered in terms of **B**, the iron bar resembles a solenoid, and considered in terms of **H**, the picture is that of a 'bar magnet' (i.e. made up of magnetic dipoles). So from this point of view, the individual 'domain' might be considered as a dipole. In the magnetized bar, the dipoles are arranged head to tail, so that only the ends are not neutralized, and from these pole layers start the lines of **H**.

8.8.2 B-H Curve of Iron

The **B**-**H** curve of a given specimen of iron can be measured experimentally. In the early days, such a plot used to be an elaborate experiment, in which **H** had to be calculated to an exact degree. For this purpose, a ring-shaped specimen would be wound uniformly with a toroidal winding, then the value of **H** inside the iron would be

$$H = \frac{NI}{2\pi r}$$

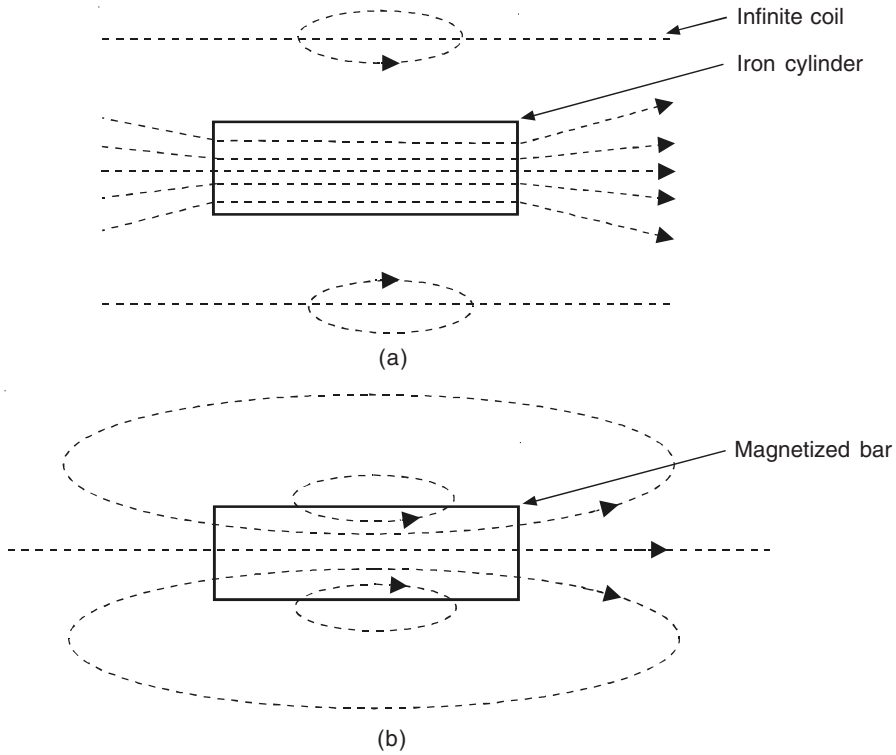


Figure 8.9 (a) **B** field of an iron cylinder inside a solenoid and (b) **B** field of the magnetized bar.

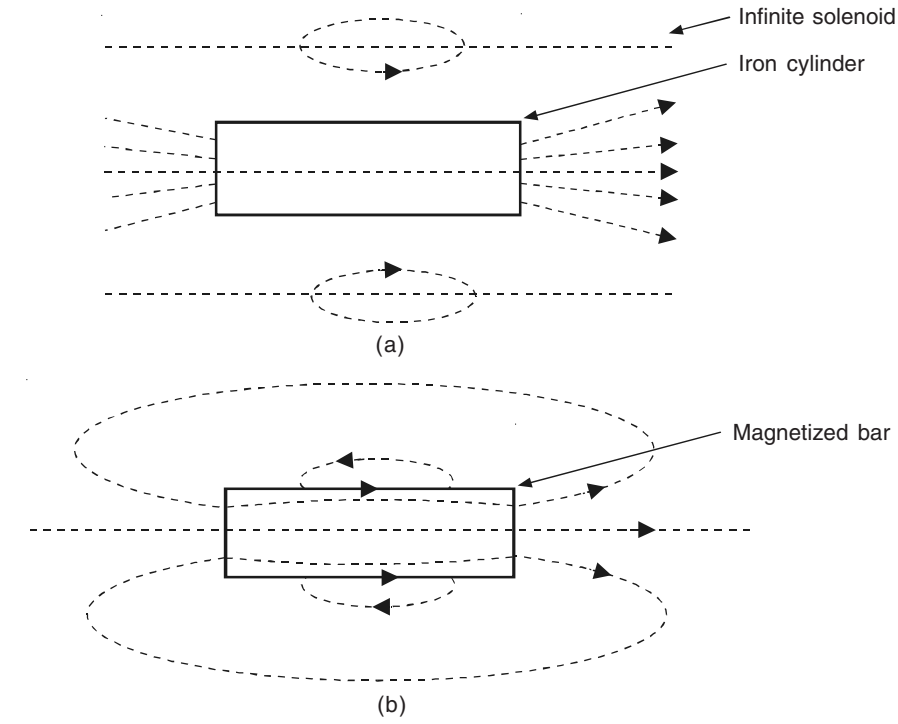


Figure 8.10 (a) **H** lines of the iron cylinder inside a solenoid and (b) **H** lines of the magnetized bar.

where I is the current per turn of the winding, and N is the total number of the turns on the ring, and r is the mean radius of the ring. **B** can be measured by using an integrating fluxmeter or a Hall-effect gaussmeter. However, nowadays the **B**-**H** plot can be obtained directly on a cathode-ray-oscilloscope screen by feeding into the X- and Y-input terminals of the scope, the signals proportional to **H**- and **B**-signals of this specimen or a specimen of any other shape.

It has to be repeated that any change in **B** is not only associated with the corresponding change in **H** but also with its history, i.e. the value of **H** that existed in the past. So we follow the **B**-**H** curve as shown in Figure 8.11.

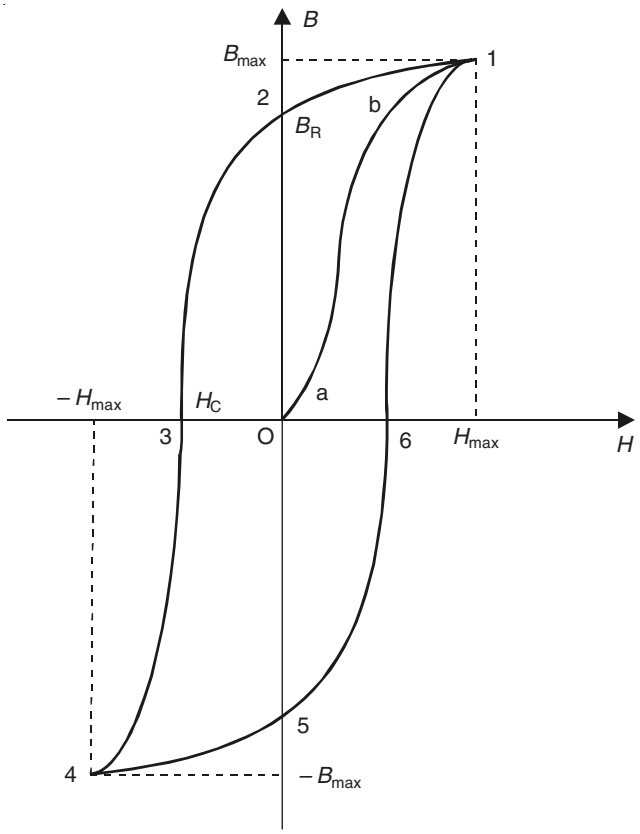


Figure 8.11 **B-H** loop for an iron specimen.

Initially, at the origin, $H = 0$, $B = 0$, $M = 0$. As the applied field is increased, there is a reversible movement of the domain walls to increase the size of the domains in the direction of the applied magnetic field at the expense of those domains whose directions are very different. This is the part O-a of the curve shown in the Figure 8.11. The second stage of the curve is the part a-b when the previously mentioned process of domain alignment continues. But at the same time, the axes of magnetization of other domains get reversed so as to get those aligned with the direction of the applied field (Figure 8.6). This part of the curve consists of a succession of small steps, and this discontinuous growth is known as the 'Barkhausen effect' as mentioned earlier. On further increase of **H** (which is done by increasing the current in the coil), a stage is reached at the value H_{max} when the axes of all the domains are aligned with the direction of **H**. The iron is then fully saturated, and this is the complete curve O-1. This part of the curve is known as the initial

magnetization curve. Any further increase in **H** will not cause a further increase in **B**. If we were to plot μ as a function of **B**, then the change in the value of μ as a function of B will be somewhat as shown in Figure 8.12. Once the saturation value is reached, the **B**-**H** curve is no longer

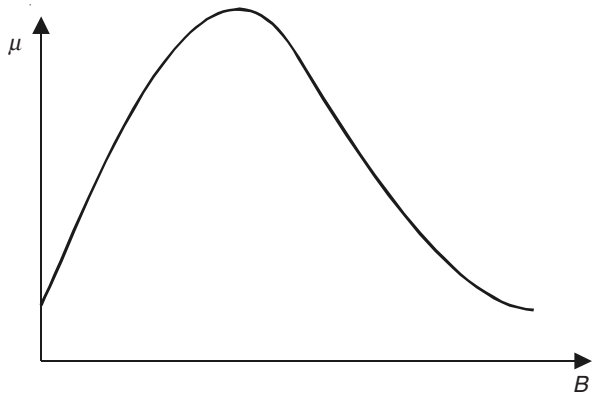


Figure 8.12 μ as a function of B for iron.

reversible. If now we start reducing H, the B values do not follow the curve 1-0, but trace a new track 1-2. When H has become equal to zero, B has a value B_R different from zero. The value of B_R is called the 'residual or remnant flux density'. The physical meaning of B_R is that due to certain effects similar to friction, the elemental current loops still remain oriented even when there is no external magnetic field, and thus produce a certain magnetic field inside the core. The remnant flux density explains the existence of the permanent magnets.

If now the current is reversed, H becomes negative, and the point traces the curve 2-3. That is to say that for a certain value H_C of H, known as the 'coercive field intensity', there will be no flux density in the core. Physically this means that the macroscopic current in the coil produces a magnetic flux density at this point which is equal in magnitude but of opposite direction to the magnetic flux density of the elemental loops which are still oriented to create a B in the positive direction. The magnetic materials with large H_C are called 'magnetically hard', and those with small H_C are called 'magnetically soft'.

If the current is increased still more in the negative direction, the point traces the curve 3-4 when H has acquired the value $-H_{max}$. If next, we decrease the current and reverse the current on reaching the zero value and then increase up to H_{max}, the point traces the curve 4-5-6-1. In fact, when **H** reaches the value $+H_{max}$, the **B** value does not exactly coincide with B_{max} at the end of the first cycle. It takes near about 10-12 cycles for the **B** to coincide with B_{max} at 1 on the initial magnetization curve. This closed curve 1-2-3-4-5-6-1 is called the 'hysteresis loop' of the material. In most of the practical applications of the ferromagnetic materials (i.e. electrical machine and transformer laminations, made up of Si-Fe alloy), we deal with sinusoidal time-harmonic macroscopic currents, and hence the hysteresis loop is possibly their most important characteristic.

It is possible that by choosing the different values of $H_{max,}$ different sizes of the hysteresis loops can be drawn for the same material, H_{max} lying at different points of the initial magnetization curve of the material, as shown in Figure 8.13. The largest loop obtained for a material, when H_{max} has reached the saturation value of **B** is the normal magnetization loop, and this loop defines the B_R and H_C of the material.

The ratio (B/H) obtained from the initial magnetization curve (also known as the normal

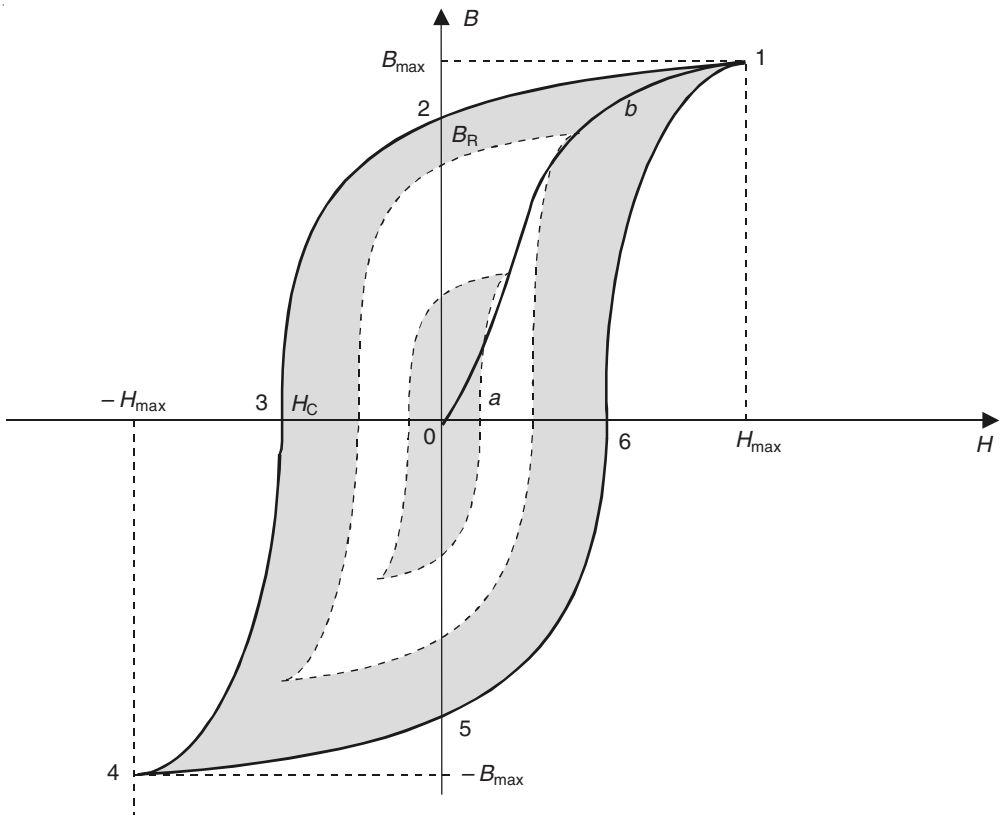


Figure 8.13 *B-H* corresponding to different values of H_{max} for a material.

magnetization curve) defines the 'normal permeability' of the material. Next, the 'differential permeability' can be defined as (dB/dH) along the normal magnetization curve at any point on it. Its value at the origin is called the 'initial permeability'. There are some practical applications in which a small alternating current has to be superimposed on a large steady magnetizing current. In these cases, the point in the *B-H* plane describes a small and very flat hysteresis loop from which the 'incremental permeability' is obtained as $\Delta B/2H_b$, where $2H_b$ is the total width of the small loop.

8.8.3 A Qualitative Explanation of the Hysteresis Loops

An in-depth, quantitative analysis of the hysteresis loops and their associated energy in the material will be discussed later. At this stage, we shall try to present a qualitative picture of this phenomenon. Using the domain theory as the basis, we can say that the process of magnetization of the ferromagnetic materials implies that the domains get rotated so as to align themselves with the external magnetizing field. This process is associated with some friction. As a result, the magnetization always lags behind the applied field which causes it (i.e. the process of magnetization). Secondly, since the orientation process is accompanied by friction, there is a certain amount of energy loss, whereby the magnetic energy gets converted into heat. These losses are known as 'hysteresis losses', and will be discussed later. This domain model also explains the behaviour of another class of magnetic materials known as 'ferrimagnetics' in which the domains

are of unequal magnitudes. They are similar to the ferromagnetics, but are non-metallic and possess high resistivities. 'Ferrites', which are mixed oxides of iron, belong to this class; and are used for high frequency transformers and antennae inside the radio receivers .

8.8.4 Temperature Dependence of Ferromagnetic Materials

It has been found that all the ferromagnetic materials lose their ferromagnetic property above a certain temperature, and above this temperature they behave like ordinary paramagnetic materials. This temperature is called the 'Curie temperature'. Hence if a permanent magnet is heated above the Curie temperature, then it loses its magnetism, and on cooling becomes completely demagnetized.

8.9 TYPES OF IRON FOR SPECIFIC MAGNETIC APPLICATIONS

Permanent magnets

We have seen earlier with reference to Figures 8.11 and 8.13, that when a specimen is magnetized to the value B_{max}, and then the magnetizing force is subsequently removed, the flux density of the specimen subsides to B_R, which is known as 'remnant flux density', and is a function of B_{max}. The greatest value of B_R is obtained by pushing up B_{max} as near the saturation region as possible. The limiting value of B_R is called the 'remnance' of the material. For good permanent magnets, this value must be as large as possible. The second requirement of the good permanent magnets is also high coercivity H_C. This value corresponds to the highest value of B_R, and again depends on pushing the B_{max} value to the saturation level as before. If a material has high coercivity, then it is obstinate in retaining its magnetization against reverse fields.

Hard carbon steels make good permanent magnets, but it has been found that certain special alloys are much better. Alnico was one of the earlier alloys developed. Its earlier composition was 53% Fe, 10% Al, 18% Ni, 13% Co, and 6% Cu. Its B_R = 0.7 Wb/m^2, and H_C = 40,000 AT/m. A number of varieties of Alnico have been developed with modified compositions, both isotropic as well as anisotropic, cast and sintered, so that B_R now ranges from 0.7 Wb/m^2 to 1.1Wb/m^2 and H_C from 40,000 AT/m to 190,000 AT/m. The drawbacks of Alnico magnets happen to be that they lack ductility and are extremely brittle. These magnets are used in measurement instruments such as voltmeters, ammeters, wattmeters, flowmeters, gyrometer astronomical instruments, electronic spectrometers, electron diffraction systems, and so on. The Curie temperatures of most Alnico varieties range from 800°C to 880°C.

Ferrites are the next category to be in wide use. At present, Barium and Strontium ferrites, Samarium-Cobalt systems (Sm-Co_5 being most widely used one), and Nd-Fe-B (Neodymium-Iron-Boron)—both isotropic and anisotropic are being used. Ferrites are used in electromechanical devices, such as electric motors (both rotary and linear), car alternators, exciters for large turbo-generators, aircraft generators, actuators for computer printers, magnetic bearings, suspensions in vehicle levitations, klystrons and magnetrons, focussing lenses in particle accelerators, and so on. Ne-Fe-B magnets are used in electric vehicles, eddy-current brakes, magnetic resonance imaging systems, and so on. Miniature magnets of this material are being used in brushless dc motors, stepper motors, dc linear motors, magnetic separators, relays and switches, audio-transducers, actuators, and so on.

Large electrical machines and transformers

In the armatures of these machines, and in the cores of large power transformers, the magnetic flux and hence the flux density B vary cyclically with time. So it is desirable to have a high permeability of the iron used so that the current required to set up the flux is as low as possible, and also a very narrow hysteresis loop so as to minimize the losses. A high saturation flux density is also desirable, in order to keep the quantity of iron to a minimum. For meeting these requirements of the alternating flux, the most suitable alloy has been found to be Si-Fe in which the Si content has been 3–4%. For the rotating machines, since the direction of the magnetic flux keeps on changing spatially, the variety of Si-Fe used in the core laminations is CRNGO (Cold rolled non-grain oriented), whereas in the transformer cores the magnetic flux reverses its direction with time only and hence CRGO (Cold rolled grain oriented) steel is used.

Magnetic shielding

For shielding instruments from the effects of stray magnetic fields, in which the flux density is usually low, it is preferable to have a material with a high initial permeability; i.e. a high μ_r at low values of **B**. One of the most suitable metals for such purposes is 'Mumetal', an alloy whose composition is 76% Ni, 17% Fe, 5% Cu, and 2% Cr. The initial permeability is 20,000 and the maximum μ_r is 80,000 as compared with 500 of silicon steels.

8.10 THE MAGNETIC CIRCUIT

As the problems involving the ferromagnetic cores have nonlinear relationships between B and H, such problems in most cases can be solved only approximately, but with a sufficient degree of accuracy. The approximate analysis resembles that of the steady electric current circuits. The basis for this analogy is the relationship in that the flux of **B** out of any closed surface is zero, and hence analogous to the circulation of the electric current in a closed circuit, as shown in Figure 8.14.

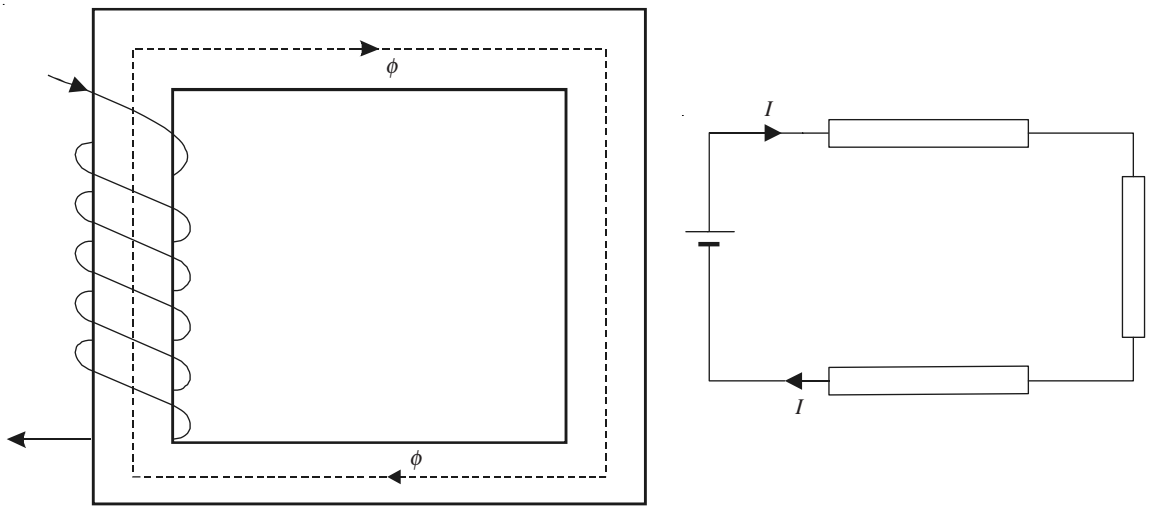


Figure 8.14 Analogy between magnetic circuit and current circuit.

In such problems, a closed magnetic path (of iron) is magnetized by a current-carrying coil. Since the iron path is a closed one, the leakage of the magnetic flux in air is negligible, and hence the total quantity of flux ϕ is same over all the cross-sections of the iron, just as in an electric circuit the current is same at all the cross-sections of the wire. The flux is the analogue of the current in the electric circuit, and the flux producing magnetizing coil is the analogue of the battery. The ampere-turns of the coil is the equivalent of the battery emf, and hence is called the magnetomotive force (mmf). Since at every point of this circuit, there is a magnetizing force **H**, we can write by using the Ampere's law,

$$\oint_C \mathbf{H} \cdot d\mathbf{l} = \Sigma I = \text{mmf} \tag{8.47}$$

comparable to

$$\int_C \mathbf{E} \cdot d\mathbf{l} = \mathcal{E} = \text{emf in the circuit.}$$

Although the *B-H* relationship is nonlinear, it is convenient to assume a linear relationship between the flux and the mmf, i.e.

$$\frac{\text{mmf}}{\text{flux}} = \text{reluctance } R_{\text{m}} \text{ of the path}$$

a quantity which is analogous to the resistance in an electric circuit.

$$\therefore \qquad R_{\text{m}} = \frac{\int \mathbf{H} \cdot d\mathbf{l}}{\phi} \tag{8.48}$$

The unit of the reluctance is ampere-turns/weber. The reciprocal of the reluctance is called the 'permeance'.

It can be assumed that H is constant in each arm of the magnetic circuit, and so $\mathbf{H} \cdot d\mathbf{l}$ can be replaced by $\sum_i H_i l_i$ for the various portions of the magnetic circuit.

$$\therefore \qquad \sum_i H_i l_i = NI \tag{8.49}$$

for the coil having N turns, and each turn carrying the current I, and l_i is the length of each portion of the magnetic path. Also, it can be assumed that

$$\phi = B \times A \tag{8.50}$$

in each limb of the circuit, where B is the flux density in the limb, and A is the cross-sectional area of the limb. When ϕ is known, B in each portion of the limb can be calculated, and hence the corresponding H can be obtained from the magnetization curve of the material, thus calculating the total ampere-turns required for the problem. When there is an air-gap in the circuit, as shown in Figure 8.15, most of the mmf would be required to maintain the flux in the air-gap, i.e.

$$\left(\frac{B}{\mu_0 \mu_{\text{r}}}\right)(l_1 + l_2) + \left(\frac{B}{\mu_0}\right) 2l_3 = \Sigma H_i l_i$$

It will be seen that even though

$$(2l_3) << (l_1 + l_2)$$

the presence of μ_{r} in the denominator of the first term on the left-hand side makes its contribution

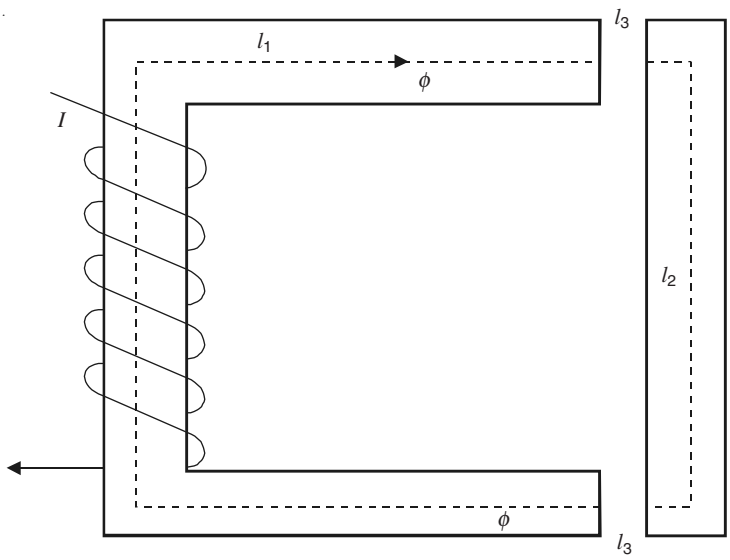


Figure 8.15 Magnetic circuit with air-gaps.

much smaller compared with the second term. A note of caution is that here we have neglected the fringing of the flux across the air-gaps as a simplifying step in the problem.

PROBLEMS

8.1 A circular iron ring of uniform permeability μ has rectangular cross-section such that its inner radius is R_i and outer radius R_o, where $R_o - R_i$ is comparable to R_i. It is wound with a uniform magnetizing winding of NI amp-turns. Show that its flux density at the mean radius $(R_o + R_i)/2$ is

$$\frac{\mu_0 \mu_r NI}{\pi(R_o + R_i)}$$

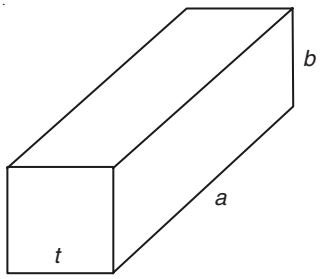
and the mean value of the flux density over the whole cross-section is

$$\frac{\mu_0 \mu_r NI}{2\pi(R_o - R_i)} \cdot \ln\left(\frac{R_o}{R_i}\right)$$

Leakage flux may be neglected.

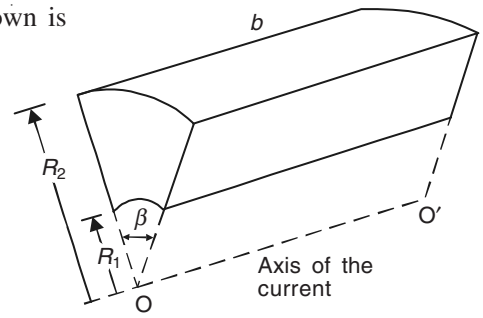
8.2 Show that the permeance of a rectangular parallelepiped region within which the flux field is parallel as shown in the adjoining figure is

$$\mathcal{P} = \mu_0 \frac{ab}{t}$$

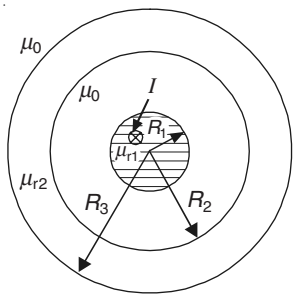


and for a sector of a rectangular toriod as shown is

$$\mathcal{P} = \mu_0 \frac{b \ln\left(\frac{R_2}{R_1}\right)}{\beta}$$



8.3 A very long conductor of circular cross-section of radius R_1 and relative permeability μ_{r1} carries a steady current I. A circular tube of relative permeability μ_{r2} and with inner and outer radii R_2 and R_3, respectively, is placed coaxially as shown in the adjoining figure. Determine the vectors **H**, **B**, **M** for all the regions.



Hint: Use Ampere's circuital law of Section 7.7, and the generalized form of the law from Section 8.4 [Eq. (8.32)] for the **M** vectors.

9 Methods of Solving Magnetostatic Field Problems

9.1 ANALYTICAL METHODS

9.1.1 Introduction

In the magnetostatic field problems, the distributions of the macroscopic currents, which are the source of the magnetic field, are usually specified in advance. If there happen to be no ferromagnetic material in the region, then the problem simply reduces to the application of the Biot–Savart's law, already stated in Eq. (7.19). Evaluation of the integral in the Biot–Savart's law can be sometimes complicated, but not impossible. Several examples of the application of this method, have been given in the Chapter 7. If the ferromagnetic bodies, with magnetic nonlinearity, are present in the region, then the exact solution of the problem becomes extremely difficult, though approximate solutions can be obtained with the aid of the computer-based methods. However, a very good insight can be obtained in a number of cases by assuming the magnetic material to be linear, with some average high value of the relative permeability ($=\mu_r$). For the analytical methods, (at least initially), we shall make this assumption.

The evaluation of the magnetic field of such a problem thus reduces to finding the **B** and **H** vectors which satisfy the specified boundary conditions of the problem. The nature of the magnetic boundary conditions have already been discussed in Chapter 8. Also, since the magnetostatic fields satisfy the same operating equations, i.e. the Laplace's and the Poisson's equations as for the electrostatic fields, the methods for solving the magnetic problems are basically the same as those described before.

9.1.2 Method of Separation of Variables

Since the basic mathematics of this method has been described in great detail in Chapter 4 (Sections 4.2 to 4.2.8) we shall not repeat the essentials of the method, but apply it directly to solve some magnetic field problems, which will highlight the application of the method to the magnetic boundaries of the problem.

9.1.2.1 Hollow cylinder in a magnetic field

This is an interesting practical problem which is a case of static shielding of the magnetic field

inside a hollow cylinder placed in a homogeneous magnetic field H_0 whose direction is normal to the axis of the cylinder as shown in Figure 9.1. We consider the problem to be two-dimensional,

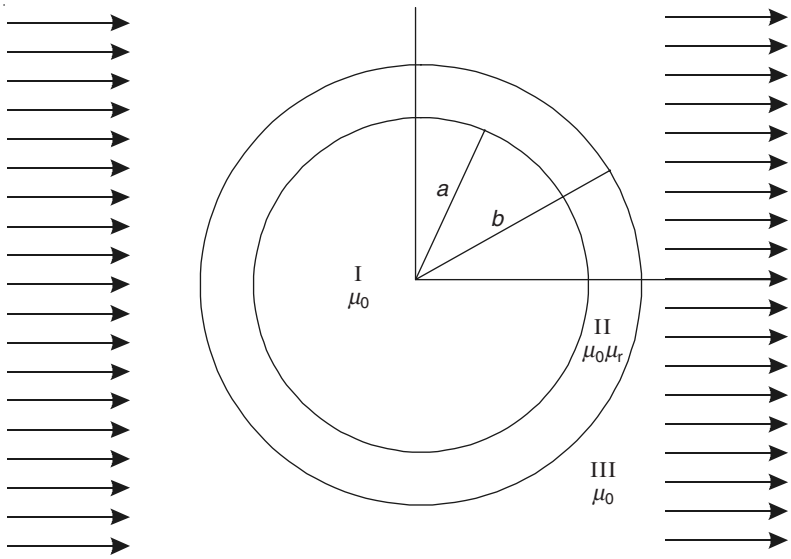


Figure 9.1 Hollow magnetic cylinder in a homogeneous magnetic field.

i.e. assume the cylinder to be infinitely long. Its inner and outer radii are a and b, respectively, and its permeability is assumed to be $\mu = \mu_0 \mu_r$.

It is a problem in the cylindrical polar coordinate system, in which there is no z-variation, i.e. the derivative $\partial/\partial z$ does not exist. Because of the fixed direction of H_0, there is no ϕ-symmetry, and the independent variables are r and ϕ. The operating equation has to be solved in the three regions:

Region I : air-space	$0 < r < a$,	$\mu = \mu_0$
Region II : iron annulus of the cylinder	$a < r < b$,	$\mu = \mu_0\mu_r$
Region III : air-space	$r > b$,	$\mu = \mu_0$

In all the three regions, $\mathbf{H} = -$ grad Ω, where Ω is the scalar magnetic potential. Since $\nabla \cdot \mathbf{B} = 0$ and $\mathbf{B} = \mu\mathbf{H}$, and since it is a two-dimensional problem, the operational equation (i.e. the Laplace's equation) is written as

$$\left(\frac{1}{r}\right)\frac{\partial}{\partial r}\left(r\frac{\partial \Omega}{\partial r}\right)+\left(\frac{1}{r^2}\right)\left(\frac{\partial^2 \Omega}{\partial \phi^2}\right)=0 \tag{9.1}$$

By the method of separation of variables, described in Chapter 4, we write Ω as

$$\Omega \equiv R\Phi \tag{9.2}$$

where $R \equiv R(r)$ and $\Phi \equiv \Phi(\phi)$.

Substituting in Eq. (9.1), we obtain the two ordinary differential equations

$$\frac{d^2\Phi}{d\phi^2} = -n^2\Phi \tag{9.3}$$

and

$$r^2 \frac{d^2R}{dr^2} + r\frac{dR}{dr} - n^2R = 0 \tag{9.4}$$

where n is the unknown separation constant, arbitrarily chosen.

The solution of Eq. (9.3) is

$$\Phi = A_n \cos n\phi + B_n \sin n\phi \tag{9.5}$$

Equation (9.4), called the Euler equation, also has the solution

$$R = C_n r^n + D_n r^{-n} \tag{9.6}$$

Hence the general solution for Ω is

$$\Omega = (A_n \cos n\phi + B_n \sin n\phi)(C_n r^n + D_n r^{-n}) \tag{9.7}$$

We consider the region I first. For this region $0 < r < a$, the solution at $r = 0$ must be finite, and so $D_n = 0$; and also the $\phi = 0$ axis (corresponding to the x-axis in Figure 9.1) has symmetry about it, and so $B_n = 0$.

$$\therefore \qquad \Omega_1 = A_{n1} r^n \cos n\phi \tag{9.8}$$

For the region II ($a < r < b$), the solution will have the more general form

$$\Omega_2 = (C_{n2} r^n + D_{n2} r^{-n}) \cos n\phi \tag{9.9}$$

For the region III ($r > b$), the field must satisfy the condition that as $r \to \infty$, Ω_3 cannot tend to zero, but must be such that $H_\infty = H_0$, i.e. $\mathbf{H}_\infty = \mathbf{i}_r H_r + \mathbf{i}_\phi H_\phi$, where

$$H_\phi = -\left(\frac{1}{r}\right)\left(\frac{\partial \Omega}{\partial \phi}\right), \quad H_r = -\frac{\partial \Omega}{\partial r} \tag{9.10}$$

$\therefore$ In the region III, the potential must be of the form

$$\Omega_3 = H_0 r \cos \phi + A_{n3} r^{-n} \cos n\phi \tag{9.11}$$

Thus there are four unknowns to be evaluated, i.e. A_{n1}, C_{n2}, D_{n2}, and A_{n3}; and there are also four conditions on the two interfaces $r = a$, and $r = b$. On $r = a$,

$$\left(\frac{\partial \Omega_1}{\partial \phi}\right)_{r=a} = \left(\frac{\partial \Omega_2}{\partial \phi}\right)_{r=a}$$

and

$$\mu_0 \left(\frac{\partial \Omega_1}{\partial \phi}\right)_{r=a} = \mu_0 \mu_r \left(\frac{\partial \Omega_2}{\partial \phi}\right)_{r=a} \tag{9.12}$$

From Eqs. (9.12) and Eqs. (9.8) and (9.9), we get

$$a^2 A_{11} - a^2 C_{12} - D_{12} = 0 \tag{9.13}$$

and

$$a^2 A_{11} - \mu_r a^2 C_{12} + \mu_r D_{12} = 0 \tag{9.14}$$

It will be seen that since there are only four constraint equations, we have only $n = 1$ to consider. The next two boundary conditions on the outer surface of the cylinder ($r = b$) are

$$\left(\frac{\partial \Omega_2}{\partial \phi}\right)_{r=b} = \left(\frac{\partial \Omega_3}{\partial \phi}\right)_{r=b}$$

and

$$\mu_0 \mu_r \left(\frac{\partial \Omega_2}{\partial \phi}\right)_{r=b} = \mu_0 \left(\frac{\partial \Omega_3}{\partial \phi}\right)_{r=b} \tag{9.15}$$

These two equations and Eqs. (9.9) and (9.11), give us

$$A_{13} - b^2 C_{12} - D_{12} = - b^2 H_0 \tag{9.16}$$

$$A_{13} + \mu_r b^2 C_{12} - \mu_r D_{12} = b^2 H_0 \tag{9.17}$$

From four simultaneous Eqs. (9.13), (9.14), (9.16), and (9.17), we evaluate the unknowns as

$$A_{11} = 2H_0 \left[\frac{2\mu_r}{(\mu_r + 1)^2 - (\mu_r - 1)^2 \left(\frac{a}{b}\right)^2}\right]$$

$$C_{12} = 2H_0 \left[\frac{\mu_r + 1}{(\mu_r + 1)^2 - (\mu_r - 1)^2 \left(\frac{a}{b}\right)^2}\right]$$

$$D_{12} = 2H_0 b^2 \left[\frac{(\mu_r - 1)\left(\frac{a}{b}\right)^2}{(\mu_r + 1)^2 - (\mu_r - 1)^2 \left(\frac{a}{b}\right)^2}\right]$$

$$A_{13} = - H_0 \left[\frac{b^2 (\mu_r^2 - 1)\left[1 - \left(\frac{a}{b}\right)\right]^2}{(\mu_r + 1)^2 - (\mu_r - 1)^2 \left(\frac{a}{b}\right)^2}\right] \tag{9.18}$$

So the potential distributions in all the three regions, i.e. Ω_1, Ω_2, and Ω_3 have been completely determined. Since our interest is in the screening effect due to the iron cylinder placed in the magnetic field, we consider the potential distribution Ω_1 in the cavity of the cylinder, i.e.

$$\Omega_1 = \left[\frac{4\mu_r}{(\mu_r + 1)^2 - (\mu_r - 1)^2 \left(\frac{a}{b}\right)^2}\right] H_0 r \cos \phi \tag{9.19}$$

Hence the magnetic field **H** or (**B** = μ_0**H**) inside the cavity is

$$H_{1\phi} = -(\nabla\Omega_1)_\phi = -\left(\frac{1}{r}\right)\left(\frac{\partial\Omega_1}{\partial\phi}\right)$$

$$= \frac{H_0 4\mu_r \sin\phi}{(\mu_r+1)^2 - (\mu_r-1)^2\left(\frac{a}{b}\right)^2}$$

$$H_{1r} = -(\nabla\Omega_1)_r = -\frac{\partial\Omega_1}{\partial r}$$

$$= -\frac{H_0 4\mu_r \cos\phi}{(\mu_r+1)^2 - (\mu_r-1)^2\left(\frac{a}{b}\right)^2} \tag{9.20}$$

$$|\mathbf{H}_1| = \sqrt{H_{1\phi}^2 + H_{1r}^2} = \frac{4\mu_r H_0}{(\mu_r+1)^2 - (\mu_r-1)^2\left(\frac{a}{b}\right)^2} \tag{9.21}$$

Hence the degree of shielding is given by the expression

$$\eta = \frac{|\mathbf{H}_1|}{H_0} = \frac{4\mu_r}{(\mu_r+1)^2 - (\mu_r-1)^2\left(\frac{a}{b}\right)^2} \tag{9.22}$$

and it depends on two factors:

1. By changing μ_r, i.e. choosing the material
2. By changing the wall thickness of the cylinder, i.e. the ratio (a/b) for a fixed value of μ_r.

The **H** field and the effects of these parameters are shown graphically in Figures 9.2 and 9.3, respectively.

We shall give further examples of the capability of this method later when the concept of the 'magnetic vector potential' is introduced in Chapter 13.

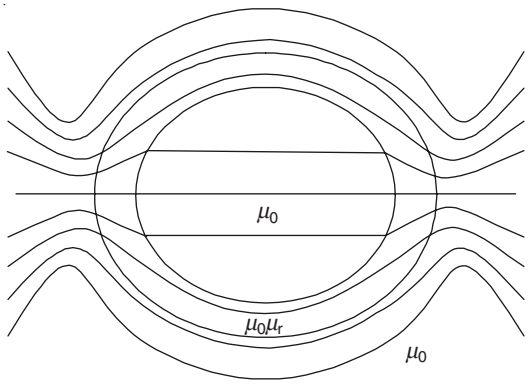


Figure 9.2 **H** field in a screening cylinder.

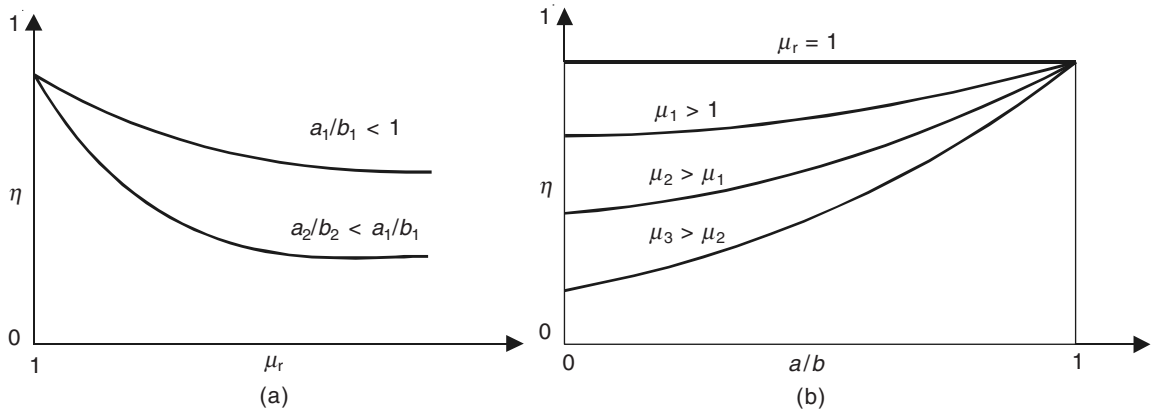


Figure 9.3 Degree of shielding η as a function of the two parameters—μ_r and a/b.

9.1.3 Conformal Transformation

As the background of this method has already been discussed in Chapter 4, Section 4.4, we shall not repeat the underlying theory of it here. The complex variable methods are equally applicable to the magnetostatic problems as the operating equations are the same as for the electrostatic fields, i.e. the Laplace's and the Poisson's equations. We shall merely illustrate the capability of such methods by solving some practical problems dealing with the magnetostatic fields.

9.1.3.1 Forces between the armature and the magnet of a contactor

We consider the armature and the magnet of a contactor as shown in Figure 9.4(a). A magnetic potential difference, established between the two parts, produces two components of force on each, the force being the consequence of the magnetic field produced in the region under consideration. The two components of the force are: (a) an attractive force between the two parts, and (b) a force tending to align the two parts, so that they become symmetrical to each other, i.e. the distances p and q change to $(p + q)/2$ on each side. For our purposes, we need to consider only one side of the configuration about the axis A A′, so that the complete configuration of Figure 9.4(a) can be synthesized from the two boundaries of Figure 9.4(b) with the appropriate lengths p and q respectively.

The first step is to evaluate the magnetic field distribution, for the configuration of Figure 9.4(b) in the z-plane which is transformed to the real axis in the t-plane of Figure 9.4(c). The corresponding points in the two planes are then:

$$\begin{aligned}
a &\rightarrow z = j\infty, & t &= -\infty & &\rightarrow a' \\
b &\rightarrow z = jg, & t &= -1 & &\rightarrow b' \\
c &\rightarrow z = -\infty, & t &= 0 & &\rightarrow c'(0) \\
d &\rightarrow z = p, & t &= \alpha & &\rightarrow d' \\
e &\rightarrow z = p - j\infty, & t &= \infty & &\rightarrow e'
\end{aligned}$$

As the vertex angles at the points jg, ∞, and p are $3\pi/2$, 0, $3\pi/2$, respectively, the Schwarz–Christoffel equation for the transformation of the z-plane to the t-plane is

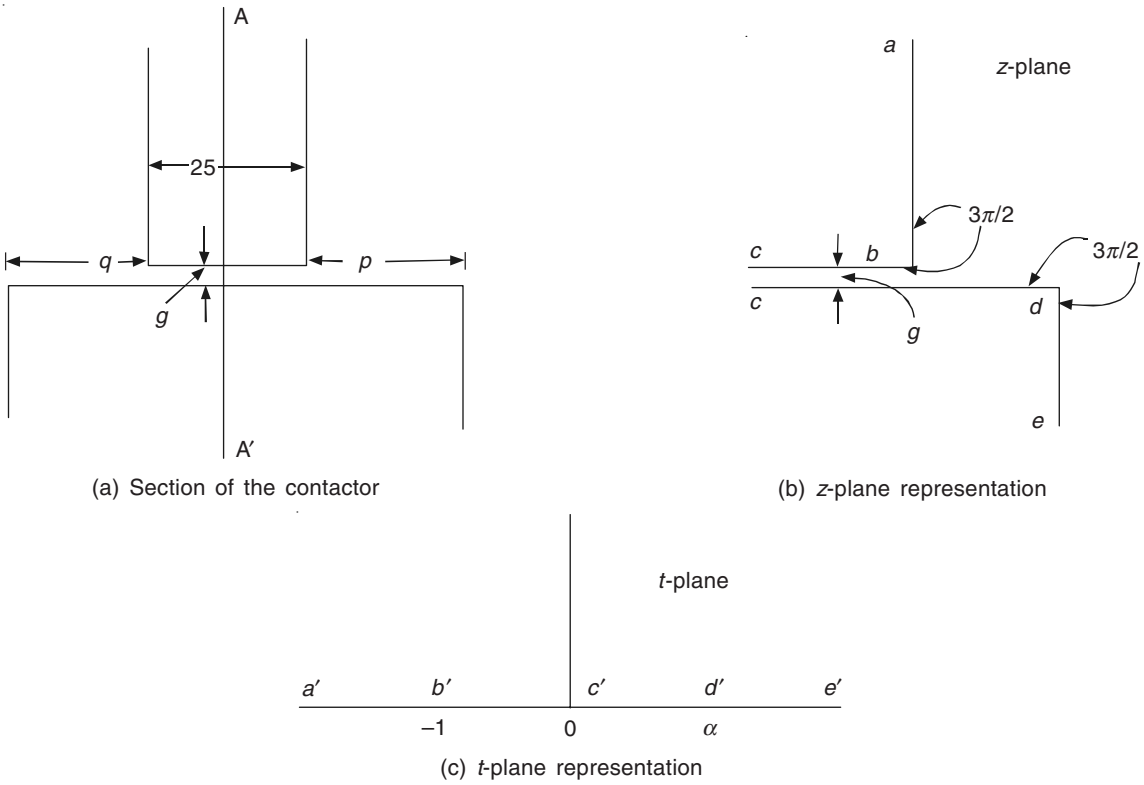


Figure 9.4 The problem of the magnet and the armature of the contactor, with the transformations.

$$\frac{dz}{dt} = K\left[\frac{\sqrt{(t+1)\,(t-\alpha)}}{t}\right] \tag{9.23}$$

To solve Eq. (9.23), we multiply its right-hand side by $\sqrt{(t-\alpha)/(t-\alpha)}$, and integrate. The result is

$$z = K\left[\frac{R(\alpha+1)}{R^2-1} + (1-\alpha)\tanh^{-1} R + j\sqrt{\alpha}\log\frac{R\sqrt{\alpha}-j}{R\sqrt{\alpha}+j}\right] + C \tag{9.24}$$

where $R = \sqrt{(t+1)/(t-\alpha)}$.

Using the calculus of residues, applied to the point $t = 0$, we get the unknown K as

$$K = -\frac{jg}{\pi\sqrt{\alpha}} \tag{9.25}$$

[This result is also obtained by direct substitution in Eq. (9.24)]. Using this value of K, and substituting for the two points $z = jg$, $t = -1$, and $z = p$, $t = \alpha$ in Eq. (9.24), the remaining unknowns come out as

$$C = 0 \qquad \text{and} \qquad \alpha = 1 + \frac{2p}{g} \pm \sqrt{\left(1+\frac{2p}{g}\right)^2 - 1} \tag{9.26}$$

Finally, if the magnetic potential difference between the two elements in the z-plane is ψ, the solution for the field in the t-plane is given by

$$w = \left(\frac{\psi}{\pi}\right) \ln t \tag{9.27}$$

and the solution for the field is obtained by eliminating t between Eqs. (9.27) and (9.24) (see Section 4.4.4.3).

From this information, the components of the force can be obtained. Since the boundaries are taken as infinitely permeable, this component of the force acts on the vertical faces, and is given by

$$F = \left(\frac{1}{2}\right) \mu_0 \int_{jg}^{j\infty} B_z^2 \, dz \tag{9.28}$$

The derivation of this expression will be given in the subsequent Chapter 11 on "Mechanical Forces and Energy Distribution in Magnetic Fields".

Equation (9.28) can be rewritten as

$$F = \left(\frac{1}{2}\right) \mu_0 \int_{jg}^{j\infty} \left(\left|\frac{dw}{dz}\right|\right)^2 dz$$

$$= \left(\frac{1}{2}\right) \mu_0 \int_{jg}^{j\infty} \left(\left|\frac{dw}{dt}\right|\right)^2 \left(\left|\frac{dt}{dz}\right|\right)^2 dz$$

By changing the independent variable to t, this expression becomes

$$F = \left(\frac{1}{2}\right) \mu_0 \int_{-1}^{-\infty} \left(\left|\frac{dw}{dt}\right|\right)^2 \left(\left|\frac{dt}{dz}\right|\right)^2 dt \tag{9.29}$$

Differentiating Eq. (9.27), we get

$$\frac{dw}{dt} = \left(\frac{\psi}{\pi}\right)\left(\frac{1}{t}\right)$$

and using Eqs. (9.23) and (9.25) along with this equation, the expression for the force becomes

$$F = \frac{\mu_0 \psi^2 \sqrt{\alpha}}{2\pi g} \int_{-1}^{-\infty} \left(\frac{dt}{t\sqrt{(t-a)\,(t+1)}}\right)$$

$$= \frac{\mu_0 \psi^2}{2\pi g} \left[\sin^{-1}\left(\frac{1-\alpha}{1+\alpha}\right) - \frac{\pi}{2}\right] \tag{9.30}$$

This is the portion of the alignment force corresponding to the right-side of Figure 9.4(a). By considering the left portion of the same figure, the resultant aligning force is obtained as

$$F = \frac{\mu_0 \psi^2}{2\pi g} \left[\sin^{-1}\left(\frac{1-\alpha}{1+\alpha}\right) - \sin^{-1}\left(\frac{1-\beta}{1+\beta}\right)\right] \tag{9.31}$$

where β corresponds to the distance q, as α corresponds to p.

9.1.3.2 Current between two infinite parallel permeable surfaces

A general form of this problem can also be solved by the method of images. At present, we consider the configuration when the current is mid-way between the permeable surfaces. The current is located at the origin of the x-y plane (complex z-plane), and the permeable surfaces are $y = \pm b$, extending to ∞ at both the ends. So the x-axis and the y-axis are both axes of symmetry, and we need to consider only (say) one side of the x-axis or one side of the y-axis or even only one quadrant. However for the solution to be of a simpler form, we choose the positive-half of the y-axis, as shown in Figure 9.5(a), and the corresponding points for the transformation from the z-plane to the t-plane are:

$$\begin{aligned} z &\to -b + j\infty \quad \to \quad t = -\infty \\ z &\to -b \quad \to \quad t = -1 \\ z &\to +b \quad \to \quad t = +1 \\ z &\to b + j\infty \quad \to \quad t = +\infty \end{aligned}$$

Hence the Schwarz–Christoffel transformation equation, corresponding to Eq. (4.160), for this problem is given by

$$\frac{dz}{dt} = \frac{S}{\sqrt{t^2 - 1}} \tag{9.32}$$

On integrating, we get

$$z = S \cosh^{-1} t + C \tag{9.33}$$

Substituting for $z = -b$, $t = -1$ and $z = +b$, $t = +1$, the unknowns are obtained as $C = b$ and $S = 2jb/\pi$.

In the z-plane, the current I sets up a potential difference of $(I/2)$ between the two halves of the boundary, meeting at the point $z = 0$, and therefore each of the t-planes must have the same potential difference, $(I/2)$, between the sections of the real axis corresponding to the two halves of

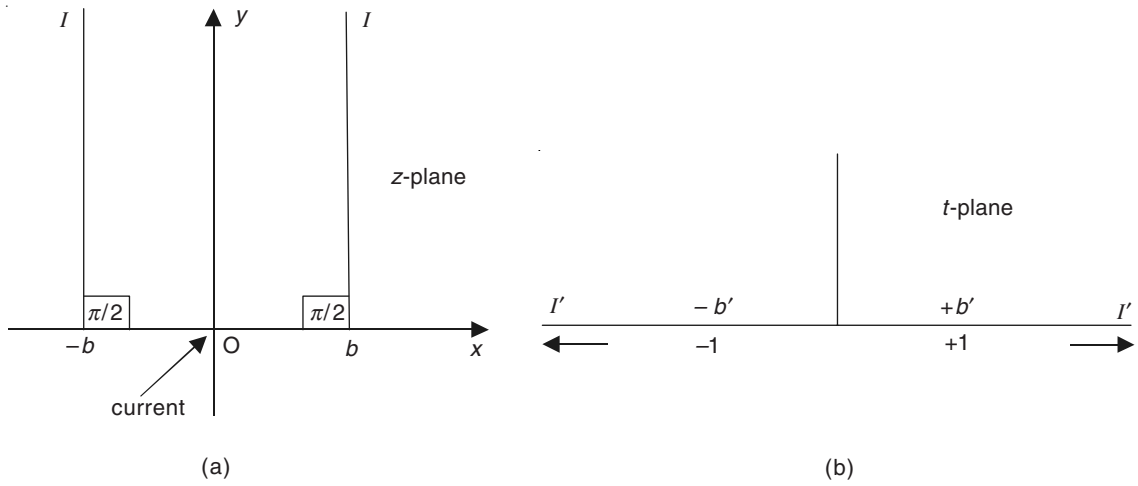


Figure 9.5 Current between two infinite parallel permeable planes: (a) z-plane representation and (b) t-plane representation.

the boundary in the z-plane. This requires the potential division at the points $t = 0$ and $t = \pm \infty$, and hence the field is given by

$$w = \left(\frac{I}{2\pi}\right) \ln t \tag{9.34}$$

Eliminating t between Eqs. (9.34) and (9.33), we get

$$w = \left(\frac{I}{2\pi}\right) \ln \left[\cosh \frac{(z-b)\pi}{2jb}\right] \tag{9.35}$$

9.1.4 Method of Images

The magnetic fields can be considered to be due to, either a distribution of steady electric currents or a distribution of magnetic dipoles. In either case, the source is a vector, as distinct from the scalar source (i.e. the electric charges) for the electrostatic fields. So the basis for the theory of images applied to the magnetostatic field problems would be the vector form of the Green's theorem (as given by Stratton) i.e. Eq. (0.59) as derived in Section 0.6.4, which is rewritten here for convenience as

$$\iiint_v \{(\nabla \times \mathbf{Q}) \cdot (\nabla \times \mathbf{P}) - \mathbf{P} \cdot (\nabla \times \nabla \times \mathbf{Q})\}\, dv = \oiint_S (\mathbf{Q} \times \nabla \times \mathbf{P}) \cdot d\mathbf{S}$$

In this equation, let us have

$\mathbf{Q} = \mathbf{P} = \mathbf{A}$, then the above equation becomes:

$$\iiint_v \{(\nabla \times \mathbf{A})^2 - \mathbf{A} \cdot (\nabla \times \nabla \times \mathbf{A})\}\, dv = \oiint_S (\mathbf{A} \times \nabla \times \mathbf{A}) \cdot d\mathbf{S} \tag{9.36}$$

Let $\mathbf{A}$ be the 'vector potential' of the magnetic flux density vector $\mathbf{B}$. (We shall discuss in detail the physical and the mathematical basis for the concept of the 'magnetic vector potential' later in Chapter 13.), i.e.

$$\nabla \times \mathbf{A} = \mathbf{B} \qquad \text{and so} \qquad (\nabla \times \nabla \times \mathbf{A}) = \nabla \times \mathbf{B} = \mu_0 \mathbf{J}$$

where $\mathbf{J}$ is the electric current density. So Eq. (9.36) becomes:

$$\iiint_v \{(|\mathbf{B}|)^2 - \mathbf{A} \cdot (\mu_0 \mathbf{J})\}\, dv = \oiint_S (\mathbf{A} \times \mathbf{B}) \cdot d\mathbf{S} \tag{9.37}$$

We compare this equation with Eq. (4.171) of Section 4.5. As in that case, the above equation states the relationship between the surface field and the volume field due to the vector potential $\mathbf{A}$. As with the scalar potential for the electrostatics, so also in this case, the uniqueness of $\mathbf{A}$ has to be proved, for the image to represent the boundary conditions correctly.

So we start by assuming that $\mathbf{A}$ has two solutions $\mathbf{A}_1$ and $\mathbf{A}_2$ both satisfying Eq. (9.37). Then both these vectors satisfy the operating equation, i.e.

$$\nabla \times \nabla \times \mathbf{A}_1 = \mu_0 \mathbf{J} \qquad \text{and} \qquad \nabla \times \nabla \times \mathbf{A}_2 = \mu_0 \mathbf{J}$$

Let $\mathbf{C} = \mathbf{A}_1 - \mathbf{A}_2$, and applying Eq. (9.36) to $\mathbf{C}$, we get

$$\iiint_v (\nabla \times \mathbf{C})^2\, dv = \oiint_S (\mathbf{C} \times \nabla \times \mathbf{C}) \cdot d\mathbf{S} \tag{9.38}$$

since $\nabla \times \nabla \times \mathbf{C} = 0$.

If the surface integral of the above equation is to be zero, then the integrand of the volume integral has to be zero.

$\therefore$ $\nabla \times \mathbf{C} = 0$, which implies that $\nabla \times \mathbf{A}_1 = \nabla \times \mathbf{A}_2$ and $\mathbf{B}_1 = \mathbf{B}_2$. The magnetic field is then uniquely defined. Then the surface integral of Eq. (9.38) can be written as

$$\oiint_S (\mathbf{C} \times \nabla \times \mathbf{C}) \cdot \mathbf{n}\, dS = \oiint_S (\nabla \times \mathbf{C}) \times \mathbf{n} \cdot \mathbf{C}\, dS$$

$$= \oiint_S (\nabla \times \mathbf{C}) \cdot (\nabla \times \mathbf{C})\, dS \tag{9.39}$$

$\therefore$ If the tangential components of $(\nabla \times \mathbf{C})$ are zero on S, then the uniqueness is guaranteed. Also, if the tangential components of $\mathbf{C}$ are constant on S, then we have

$$\oiint_S (\nabla \times \mathbf{C}) \times \mathbf{n} \cdot \mathbf{C} dS = \mathbf{C} \cdot \oiint_S (\nabla \times \mathbf{C}) \times \mathbf{n}\, dS \tag{9.40}$$

But

$$\oiint_S (\nabla \times \mathbf{C}) \times \mathbf{n} dS = - \iiint_v (\nabla \times \nabla \times \mathbf{C})\, dv = 0 \tag{9.41}$$

$\therefore$ The constant tangential components of $\mathbf{C}$ also assure uniqueness. Thus we have the conclusion that the criterion for the uniqueness is that either the tangential components of $\mathbf{C}$ shall be constant or the tangential components of $(\nabla \times \mathbf{C})$ shall be zero everywhere on S. In terms of the magnetic flux density $\mathbf{B}$, this means that either normal $\mathbf{B}$ $(=\mathbf{B}_n)$ or the tangential $\mathbf{B}$ $(=\mathbf{B}_t)$ must be specified everywhere on S. So now we can formulate the image problem of the magnetostatic fields as follows.

We consider a closed surface S enclosing a volume v. We then determine the normal magnetic field (or the tangential magnetic field) on S, caused by the induced currents on the boundaries. Then we find a distribution of currents outside S, so as to produce on S the same normal or tangential $\mathbf{B}$. Then this current distribution gives within v the same magnetic field as is given by the induced currents on the boundaries. This is the required image distribution. As in the electrostatic case, the images must lie outside v, and so must be virtual.

Alternatively, the magnetostatic field can be considered to have its sources to be made up of the magnetic dipoles. All the currents then can be replaced by the equivalent magnetic shells. The magnetic field can then be derived from a scalar potential as has been done for the electrostatic field in Section 4.5. Thus the formulation of the problem is identical whether we start from the currents or the dipoles.

9.1.4.1 Line current parallel to the face of a semi-infinite magnetic slab and its generalization

This was one of the earliest problems in magnetic field images solved by Dr. G.F.C. Searle in 1898, in which he considered the configurations as shown in Figures 9.6(a) and (b). However, we shall solve a generalized problem and then show that the sets of Searle's results are the special cases of this general problem.

The problem we consider is as shown in Figure 9.7, where we have the plane interface between the media 1 and 2 of permeabilities μ_1 $(= \mu_0 \mu_{r1})$ and μ_2 $(= \mu_0 \mu_{r2})$ respectively. The magnetic field is produced by a real conductor at O, carrying a current I, located in the region 1.

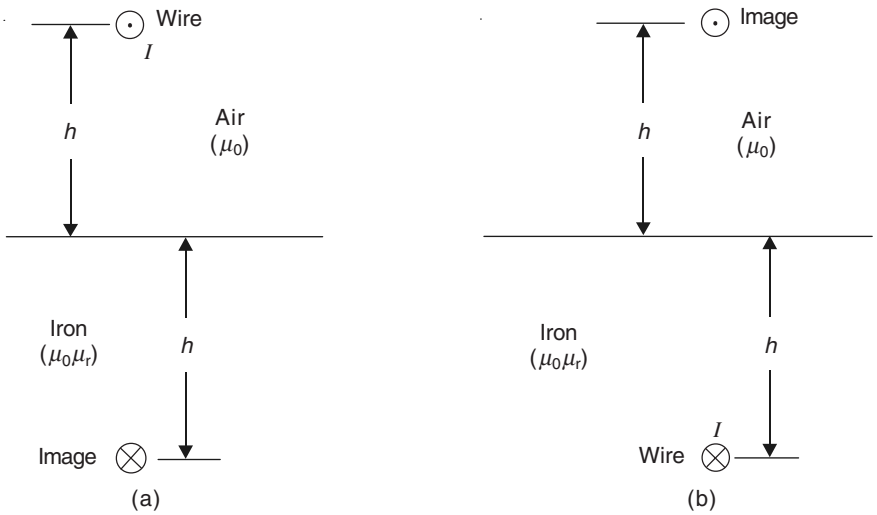


Figure 9.6 Images of a wire carrying a current *I:* (a) outside a plane block of iron of permeability $\mu_0\mu_r$ and (b) embedded in the same block of iron.

To find the effect of the medium 2 on the resultant field in 1, we place an 'image conductor', carrying a virtual current αI at the point of the optical image of I at O, i.e. at O′ as shown in Figure 9.7, replacing the region 2 by the region 1.

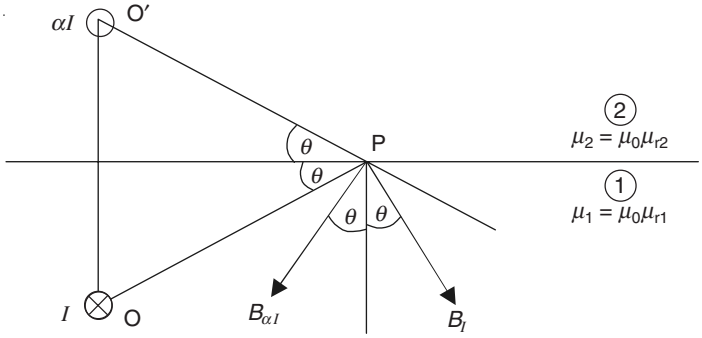


Figure 9.7 Electromagnetic image of a current-carrying conductor in a general problem.

So now we have the magnetic flux density at the point P on the interface between the two regions, which looking from the region 1 of permeability μ_1 is

$$B_I = \frac{\mu_1 I}{2\pi r} \quad \text{and} \quad B_{\alpha I} = \frac{\mu_1 \alpha I}{2\pi r}$$

∴ The resultant normal component of **B** at P on the interface is

$$B_{n1} = B_I \cos\theta + B_{\alpha I} \cos\theta = \left(\frac{\mu_1 I}{2\pi r}\right)(1 + \alpha)\cos\theta \tag{9.42}$$

and the resultant tangential component of **H** at P is

$$H_{t1} = \frac{1}{\mu_1} B_I \sin\theta - \frac{1}{\mu_1} B_{\alpha I} \sin\theta = \frac{I}{2\pi r}(1 - \alpha)\sin\theta \tag{9.43}$$

If now, we consider P from the side of the region 2, we replace the real current I at O by an imaginary current $(1 + \beta)I$ at the same point O, with the whole region of permeability μ_2, then at P

$$B_{(1+\beta)I} = \frac{\mu_2 (1+\beta) I}{2\pi r}$$

$\therefore$ The resultant normal component of **B** at P now is

$$B_{n2} = \left[\frac{\mu_2 (1+\beta) I}{2\pi r} \right] \cos\theta \tag{9.44}$$

and the resultant tangential component of **H** at P is

$$H_{t2} = \frac{B_{(1+\beta)I}}{\mu_2} \sin\theta = \left[\frac{(1+\beta) I}{2\pi r} \right] \sin\theta \tag{9.45}$$

Now applying the conditions on the interface, we have

$$B_{n1} = B_{n2} \qquad \text{and} \qquad H_{t1} = H_{t2} \tag{9.46}$$

We thus get

$$\mu_1(1 + \alpha) = \mu_2(1 + \beta) \qquad \text{and} \qquad 1 - \alpha = 1 + \beta$$

$\therefore$

$$\alpha = \frac{\mu_2 - \mu_1}{\mu_2 + \mu_1} \quad \text{and} \quad \beta = \frac{\mu_1 - \mu_2}{\mu_2 + \mu_1} \tag{9.47}$$

Applying Eq. (9.47) to Searle's problem, i.e. first, let us assume μ_1 to be in iron $(=\mu)$ and μ_2 to be in air = 1.

So this is the case of the current-carrying conductor embedded in iron, which is Searle's problem in Figure 9.6(b). It states that in iron, the field at P is due to:

I at O, the source point, and $(1 - \mu)/(1 + \mu)\ I$ at O′—the image point, i.e.

$(\mu - 1)/(\mu + 1)I$ at O′ is in the opposite direction to I at O (since $\mu > 1$).

In the air the field at P is due to:

$2\mu/(\mu + 1)I$ at O in the same direction as the real source current I.

Next we let μ_1 to be in air (=1) and μ_2 to be in iron $(=\mu)$, then we have the current-carrying conductor located in air, parallel to the magnetic medium interface, which is Searle's problem in Figure 9.6(a).

Then, in air, we have the field at P as due to:

I at O, which is the source point in air, and $(\mu - 1)/(\mu + 1)I$ at O′, the image point. In this case, the image current is in the same direction as the source current at O.

And, in iron, the field at P is due to: $2\mu/(\mu + 1)I$ at O with its image current in the same direction as the source current I.

9.1.4.2 *Line current inside the circular cylinder, and parallel to the axis of the cylinder*

The source current is I at the point A inside the cylinder, flowing into the plane of the paper, parallel to the axis of the cylinder. We consider the magnetic field at a point P on the interface circle of the two media of permeabilities—μ_1 $(= \mu_0 \mu_{r1})$ and μ_2 $(= \mu_0 \mu_{r2})$—as shown in Figure 9.8.

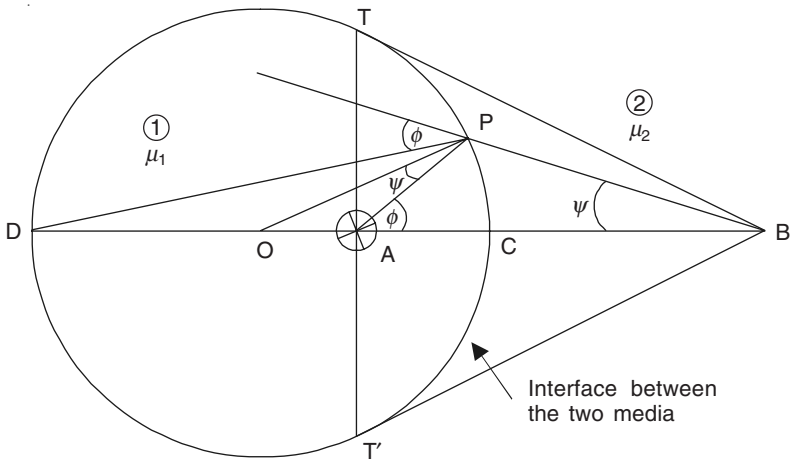


Figure 9.8 Current-carrying conductor inside the cylinder.

Hence, we have that:

1. the magnetic field in the cylinder of permeability μ_1 is due to the source current I at A and an image current αI at its inverse point B (for the derivation of the inverse point of a circle, see Appendix 2).
2. the magnetic field outside the cylinder, where the permeability is μ_2 is due to an image current βI at the (source) point A, and another image current $(1 + \gamma)I$ at O, the centre of the circle under consideration.

Mathematically, the magnetic field at P (= **B**) is:

In Region 1:

$$\frac{I}{2\pi \text{AP}} \text{ (due to } I \text{ at A, and normal to AP)} + \frac{\alpha I}{2\pi \text{BP}} \text{ (due to } \alpha I \text{ at B, and normal to BP)}$$

In Region 2:

$$\frac{\beta I}{2\pi \text{AP}} \text{ (due to } \beta I \text{ at A, and normal to AP)} + \frac{(1+\gamma)I}{2\pi \text{OP}} \text{ (due to } (1 + \gamma)I \text{ at O, and normal to OP)}$$

So to evaluate the coefficients α, β, and γ, we apply the two boundary conditions at the point P on the interface between the two media.

Hence, for the continuity of $\mathbf{B}_n$, the **B** components normal to the circle, i.e. along the radius OP at P, are:

$$\left[\left(\frac{\mu_1 I}{2\pi \text{AP}}\right)\sin\psi, \left(\frac{\mu_1 \alpha I}{2\pi \text{BP}}\right)\sin(\pi - \phi)\right] \text{ in region 1, and}$$

$$\left[\left(\frac{\mu_2 \beta I}{2\pi \text{AP}}\right)\sin\psi, 0\right] \text{ in region 2}$$

due to currents I, αI, βI, and $(1 + \gamma)I$, respectively.

$$\therefore \qquad B_{n1} = B_{n2} \Rightarrow \frac{\mu_1 I \sin\psi}{\text{AP}} + \frac{\mu_1 \alpha I \sin\phi}{\text{BP}} = \frac{\mu_2 \beta I \sin\psi}{\text{AP}}$$

Since AP sin ϕ = BP sin ψ,

$$\therefore \qquad \mu_1(1 + \alpha) = \mu_2\beta \tag{9.48}$$

Next we consider the **H** field components tangential to the circle, i.e. at P the components normal to OP are:

$$\left[\left(\frac{I\cos\psi}{2\pi\text{AP}}\right), \frac{\alpha I\cos(\mu-\phi)}{2\pi\text{BP}}\right] \text{ in region 1, and}$$

$$\left[\left(\frac{\beta I\cos\psi}{2\pi\text{AP}}\right), \left(\frac{(1+\gamma)I}{2\pi\text{OP}}\right)\right] \text{ in region 2}$$

due to currents I, αI, βI, and $(1 + \gamma)I$, respectively.

$$\therefore \qquad H_{t1} = H_{t2} \Rightarrow \frac{I\cos\psi}{2\pi\text{AP}} - \frac{\alpha I\cos\phi}{2\pi\text{BP}} = \frac{\beta I\cos\psi}{2\pi\text{AP}} + \frac{(1+\gamma)I}{2\pi\text{OP}}$$

or

$$\cos\psi - \alpha\left(\frac{\sin\psi}{\sin\phi}\right)\cos\phi = \beta\cos\psi + (1+\gamma)\frac{\sin(\phi-\psi)}{\sin\phi} \tag{9.49}$$

Since there are three unknowns, we need one more constraint, and so we take the line integral over the whole circle, and we get

$$I = \beta I + (1 + \gamma)I \tag{9.50}$$

$$\therefore \qquad \beta = -\gamma \tag{9.51}$$

Combining Eqs. (9.51) and (9.49), we get

$$\beta = 1 - \alpha \tag{9.52}$$

and with Eq. (9.48), the coefficients are obtained as

$$\alpha = \frac{\mu_2 - \mu_1}{\mu_2 + \mu_1}, \qquad \beta = \frac{2\mu_1}{\mu_2 + \mu_1}, \qquad \gamma = \frac{2\mu_1}{\mu_2 + \mu_1} \qquad \text{and} \qquad (1+\gamma) = \frac{\mu_2 - \mu_1}{\mu_2 + \mu_1} \tag{9.53}$$

Note: For the relationships between the angles, refer to Appendix 2.

9.1.4.3 Two limiting cases of the cylinder problem

First case. The cylinder is of iron, i.e. $\mu_{r1} = \mu_r$, and the conductor carrying the current I is embedded in it; the outside is air, i.e. $\mu_{r2} = 1$. Hence the results for this case can be summarized as follows:

The magnetic field in the iron (i.e. in the cylinder) is due to *the current I at* A and an image current $(1 - \mu)/(1 + \mu)I$ at its inverse point B, i.e. $-[(\mu - 1)/(\mu + 1)]I$ *at the inverse point* B *of opposite sign to I at* A.

The magnetic field in the air (i.e. outside the iron cylinder) is due to *an image current* $[2\mu/(\mu + 1)]I$ *at* A, *in the same direction* as the source current I at A, and another image current $(1 - \mu)/(1 + \mu)I$ at O or $(\mu - 1)/(\mu + 1)I$ *at* O *of opposite sign to I at* A.

Second case. We reverse the two media characteristics, i.e. the outside is iron with $\mu_{r2} = \mu_r$, and the cylinder is air, i.e. it is a cylindrical hole in the iron medium ($\mu_{r1} = 1$ for the air medium), and

the current-carrying conductor is now located in the hole, carrying the current I and is parallel to the axis of the hole.

Now the magnetic field in the air is due to the *source current I at* A, and an *image current* $(\mu - 1)/(\mu + 1)I$ *at the inverse point* B, *in the same direction as I.*

The magnetic field in the iron is due to an *image current* $[2/(\mu + 1)]I$ *at* A *in the same direction as the source current I*, and a second *image current* $(\mu - 1)/(\mu + 1)I$ *at the centre* O *in the same direction as I.* Similarly, we can solve the problem of the source current located outside the cylinder.

(*Note:* Compare the Milne–Thompson circle theorem in Fluid dynamics, Ref.: *A Text Book of Fluid Dynamics*, by F. Chorlton.)

9.1.4.4 Images of current loops and current elements

So far we have considered only the infinitely long current-carrying conductors whose images on a plane interface are shown in the Figure 9.9(a) and (b). From these images, the images of the current loops can be obtained directly. Figure 9.10 shows three circular loops located near an interface between two media of different permeabilities. The first loop with its plane coincident with the plane of the paper and normal to the interface surface, the second one with its plane normal to the plane of the paper as well as to the interface surface, whilst the third one having its

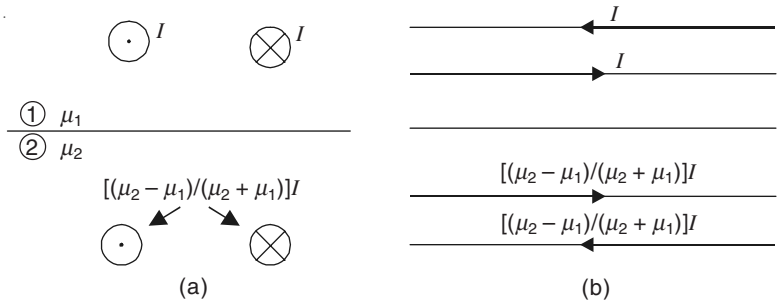


Figure 9.9 Images of infinite line currents on a plane interface.

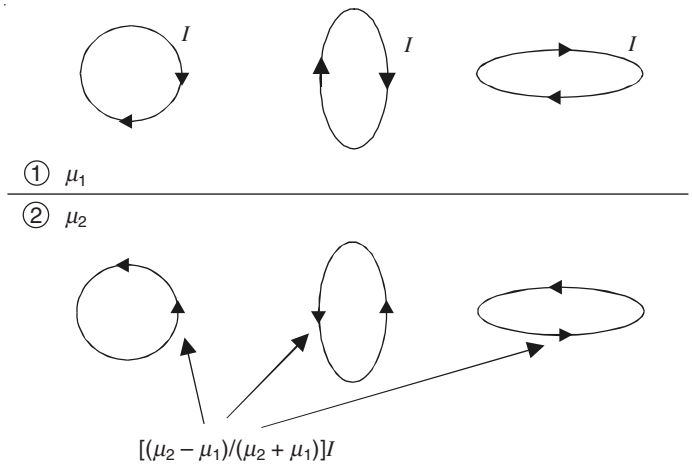


Figure 9.10 Images of current loops.

plane normal to the plane of the paper but parallel to the interface surface, are shown with their corresponding images in the interface surface.

From these loop images, it would seem that one could state a rule for the images of the 'current elements' as shown in Figure 9.11. Such a rule was stated first by Schelkunoff in 1943,

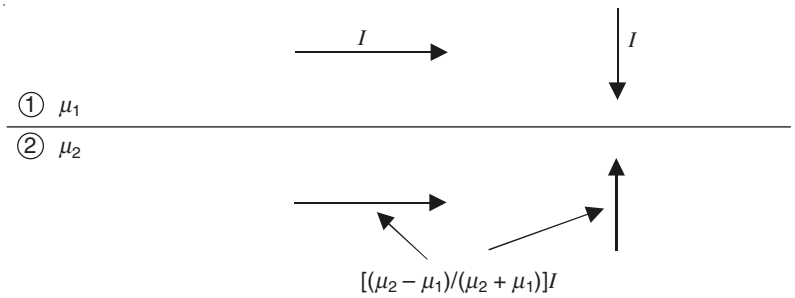


Figure 9.11 Supposed images of current elements.

particularly for the conducting sheets. But a consideration of the Biot–Savart's law shows that this is not correct, particularly for the elements normal to the interface, because such an element in isolation will not be affected by the other medium. Also, since we have not studied the time-varying currents so far, and their effects on the media of finite conductivity [in fact, up to this stage, we have restricted our discussion to the media of different magnetic permeabilities with zero electrical conductivity or infinite electrical resistivity ($\rho \to \infty$ or $\sigma \to 0$), and infinite electrical conductivity ($\sigma \to \infty$)], we shall leave this point for a discussion later. However we shall mention this much at this stage (which shall be proved later in this book) that whilst the steady current elements are a physical impossibility, it is possible to have physically, what is called the *electric doublets* (or *oscillating dipoles*) which are alternating current elements with time-varying charges at the ends of the elements. The images of these elements in iron (of finite permeability and zero electrical conductivity) would be the same as for the steady current elements. However, when a medium of finite electrical conductivity is brought in, the situation changes significantly. This is because these time-varying currents induce currents in the medium of finite electrical conductivity. In particular, if a sheet of infinite conductivity is brought in the vicinity of such doublets, what happens is that a current sheet is induced in the surface of the sheet, causing all the magnetic flux density to be parallel to such a surface, and zero normal **B**. This amounts to saying that the surface has 'effective zero permeability'. An in-depth analysis of such a phenomenon will be found later in this book. We shall now show the images of the doublets (oscillating dipoles) and the current loops in an infinitely conducting sheet and an infinitely permeable material [Figures 9.12(a) and (b)].

Finally, the images, when the source currents are embedded in iron, are shown in Figure 9.13(a) with the iron having the relative permeability μ_r; and when μ_r tends to infinity the corresponding images are shown in Figure 9.13(b).

9.1.4.5 The method of images applied to the end-windings of electrical machines (The case of partially embedded coils)

The calculation of the mechanical forces on the machine end-windings in large rotating machines is an important requirement from design considerations. The method of images has been used for

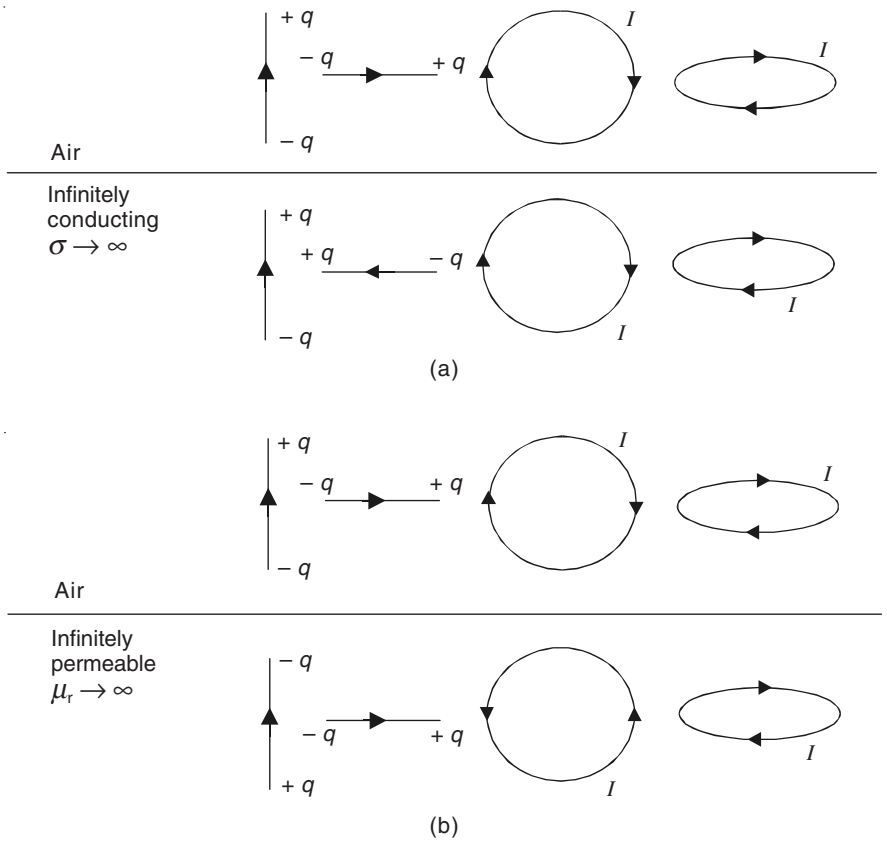


Figure 9.12 Images of doublets and current loops in the two limiting boundaries: (a) infinitely conducting and (b) infinitely permeable.

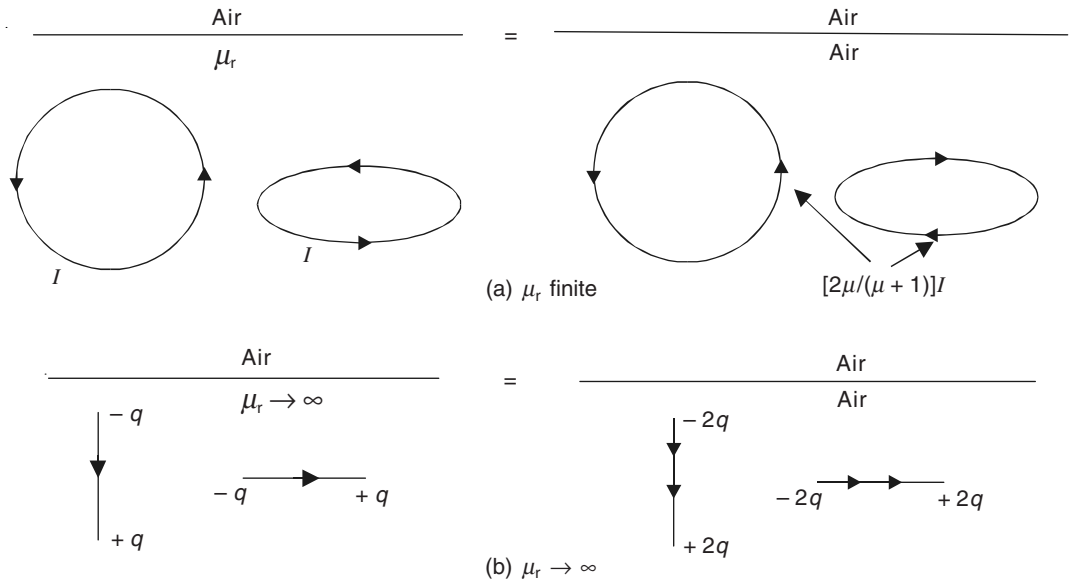


Figure 9.13 Images when the sources are embedded in iron of (a) finite μ and (b) infinite μ.

making the calculations easier by enabling the designer to take account of the effects of the boundaries on the geometry of the end-windings. The most important boundary is the core-end-surface to which the coil-ends are perpendicular. The core-end-surface is either magnetic (sometimes being infinitely permeable) or at times infinitely conducting (being a copper sheet). The coil-ends are taken as partially embedded in the core-end-surface, though in reality they pass through the slots in the core; and furthermore there is an air-gap between the stator-core and the rotor-core. The result is that the coil-ends can be considered as partially embedded coils with an air-gap discontinuity between the core-end-surfaces of the stator and the rotor of the machine. So we solve the problem in two stages: (1) the partially embedded coils with a continuous core-end-surface and (2) the air-gap discontinuity added to the oil-ends.

Stage 1. *The image of a partially embedded coil-end* on the basis of the current element image (as was used in 1950s and early 1960s) gives incorrect result, because it produces current discontinuities on the interface boundary, as is obvious from Figure 9.14.

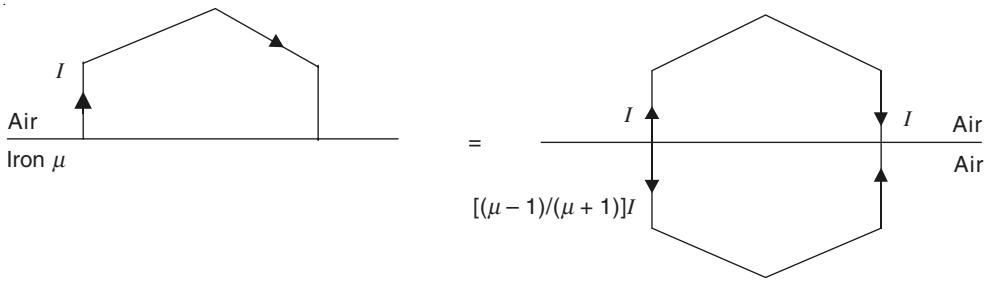


Figure 9.14 Incorrect image of the partially embedded coil, in the air-region.

In fact, the representation of the problem itself is incorrect, because the coil, though partially embedded, does have the conductors in the iron region, carrying the same current. The correct image was suggested by both Hammond and Carpenter in slightly different manners, both producing the correct effects. The images in air are shown in Figure 9.15, and the corresponding images giving the field in iron are shown in Figure 9.16.

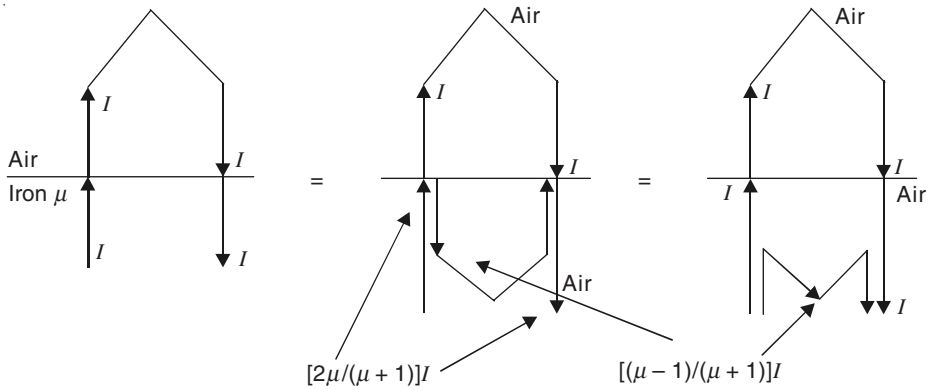


Figure 9.15 Alternative images in air, for the partially embedded coil-ends as derived by Hammond and Carpenter respectively.

The image in the perfectly conducting surface would be such that all the magnetic field in air would be tangential to the surface and the normal **B** would be zero, and the image of the configuration would obviously be as shown in Figure 9.17.

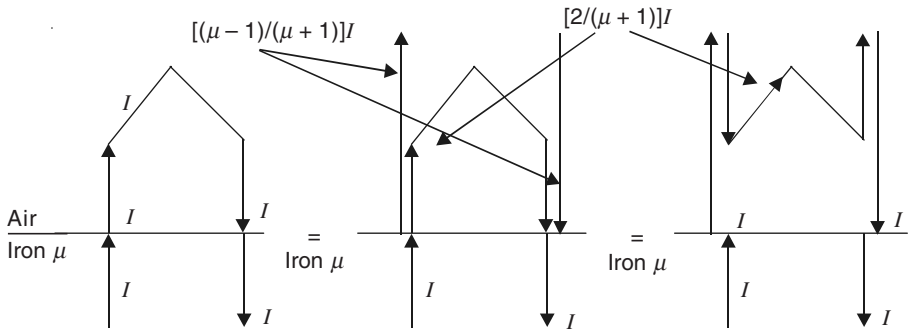


Figure 9.16 Images of partially embedded coil-end, giving the field in iron, as obtained by Hammond and Carpenter respectively.

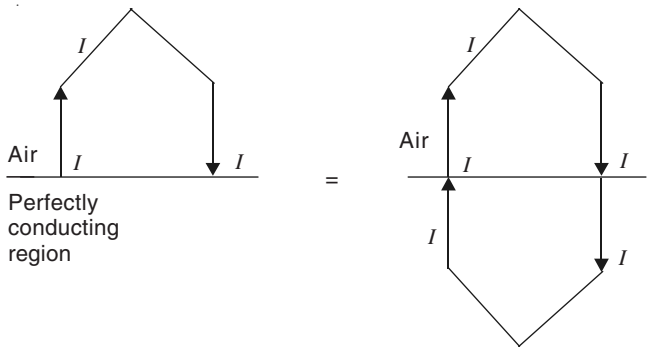


Figure 9.17 Image of a coil-end embedded partially in a perfectly conducting region, for the air-region.

Stage 2. (*The image for an air-gap discontinuity*) In the rotating machines, in the end-region, the coil-ends in reality do not project out from a continuous block of iron, but from two blocks, separated by a small air-gap, which has a drastic effect on the magnetic field near the iron, particularly when its permeability is quite large. At this discontinuity, in its immediate vicinity the tangential flux disappears and the m.m.f. appears as a potential difference between the two sides of the gap.

We shall look at the problem, initially considering a single conductor projecting out of the air-gap between the two infinitely permeable iron blocks, and then use the principle of superposition to represent the coil-end itself, noting however that the shape of the overhang of the coil does not affect the magnetic flux across the air-gap. The P.D. across the air-gap, mentioned above, results in producing the fringing flux, as shown in Figure 9.18(a). The circuit equivalent of the air-gap and the image of the complete system are shown in Figure 9.18(b) and (c).

Next, we use the principle of superposition to represent a complete coil-end, in which case the two coil-sides are effectively superposed with a displacement between them, equal to the span of the coil, remembering that the shape of the overhang portion does not affect the air-gap fringing flux. This is shown in Figure 9.19. However this image does not represent correctly the air-gap in a real machine. This is because it is implicit in the above representation that the air-gap extends to infinity on the outsides of the two coil-sides, which of course is not correct for a real machine. The air-gap around the periphery of the rotor is of finite length and hence, only if the machine has two poles and the coils are full-pitched (i.e. spanning one complete pole-width), then

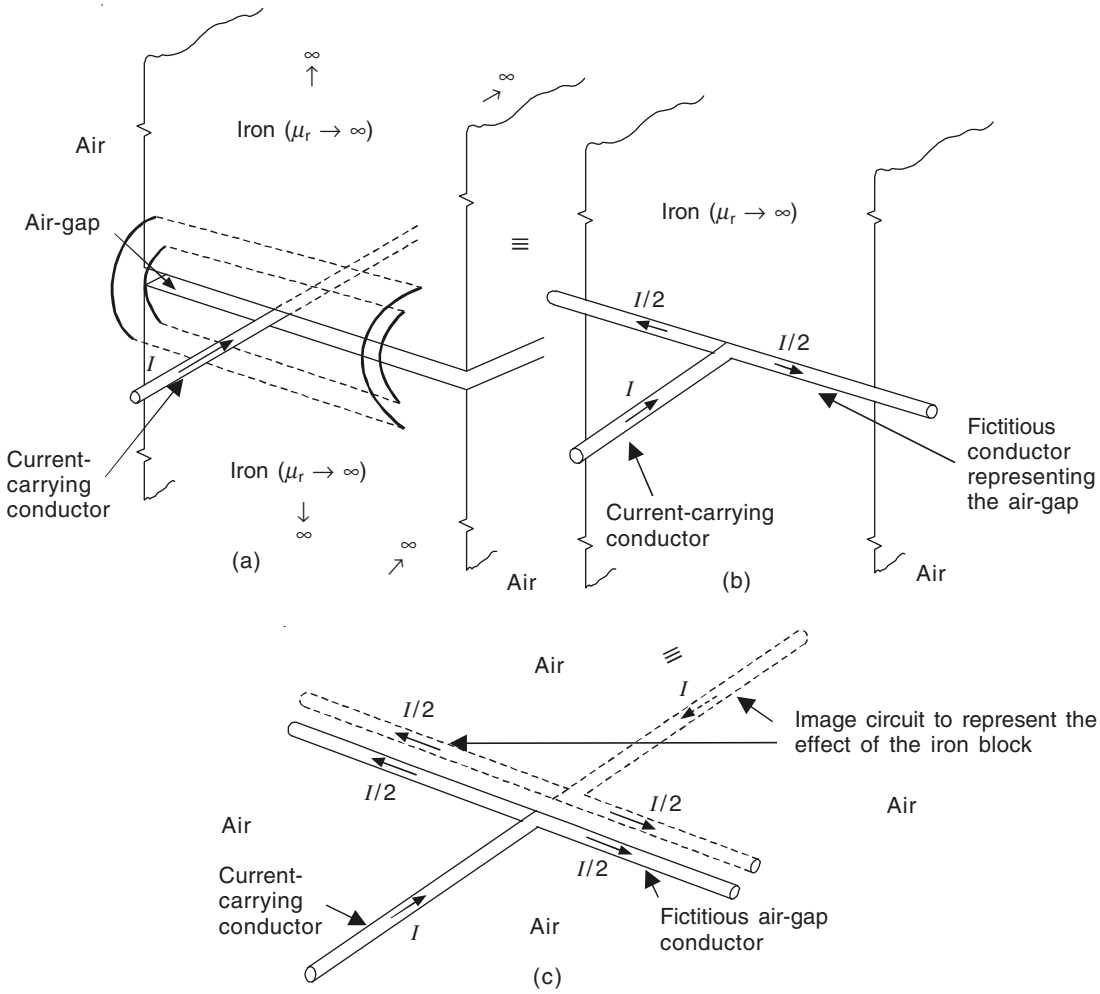


Figure 9.18 A line conductor penetrating an infinitely permeable iron block with a short width air-gap. (The block is semi-infinite, and the air-gap and the conductor extend to infinity. The diagram shows a section of the whole region.) (a) Current-carrying conductor in the air-gap. It also gives a pictorial view of the imaginary fringing flux across the air-gap. (b) Air-gap discontinuity from the iron has been replaced by its fictitious current-carrying conductor. (c) Iron has been removed and its effect produced by the image circuit.

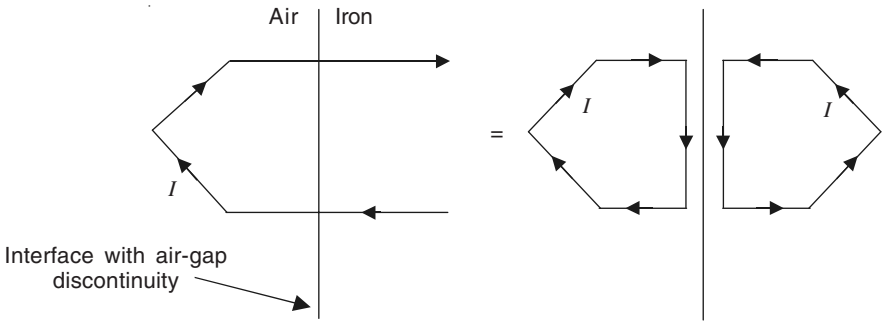


Figure 9.19 Image system for a coil-end in an infinitely long air-gap.

only the fictitious conductor representing the air-gap would break up into two parts, as shown in Figure 9.20, of equal magnitude. This is because as the air-gap is of finite length, there would be

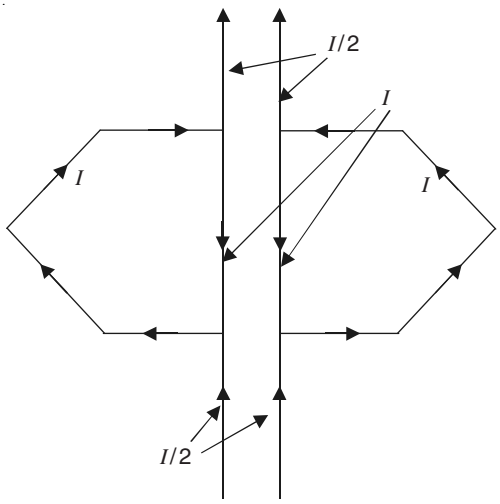


Figure 9.20 Image of a full-pitched coil-end in an infinitely permeable iron-block with finite-length air-gap.

equal fringing flux across the gap, both inside and outside the span of the coil-end, the permeance of both the parts being equal.

However not all machines have only two poles and the coils are not mostly full-pitched. Thus in these multipole machines (i.e. 4, 6, 8, ... or much greater), the reluctance of the two paths (i.e. inside a coil-span and outside it) would be unequal, and hence the fictitious current in the air-gap image conductor would be divided unequally as shown in Figure 9.21, with the current break-up into two parts given by the equations:

$$I_1 + I_2 = 2I, \qquad I \text{ being the conductor current} \tag{9.54}$$

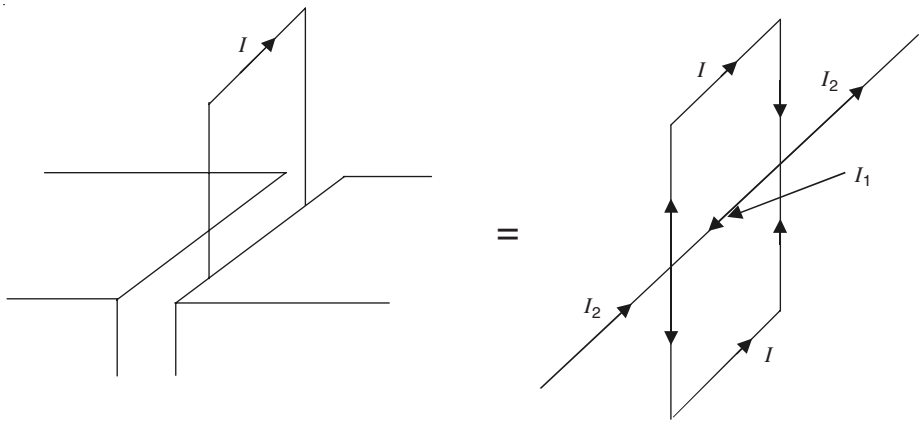


Figure 9.21 Image representation of air-gap of finite length and a coil-end of any general pitch.

and

$$I_1 \times \frac{2\pi}{p} = I_2 \times \left(2\pi - \frac{2\pi}{p}\right) \tag{9.55}$$

where p is the number of poles in the machine (also assuming the coil to be full-pitched). Hence,

$$I_1 = \frac{2(p-1)}{p} \quad \text{and} \quad I_2 = \left(\frac{2}{p}\right) I \tag{9.56}$$

If the coil is short-pitched by an angle β, instead of being full-pitched, then Eq. (9.55) is replaced by

$$I_1^{\mathrm{S}} \times \left(\frac{2\pi}{p} - \beta\right) = I_2^{\mathrm{S}} \times \left[2\pi - \left(\frac{2\pi}{p} - \beta\right)\right] \tag{9.57}$$

and

$$I_1^{\mathrm{S}} = \left[\frac{2\left(p - 1 + \dfrac{\beta p}{2\pi}\right)}{p}\right] I \quad \text{and} \quad I_2^{\mathrm{S}} = \left[\frac{2\left(1 - \dfrac{\beta p}{2\pi}\right)}{p}\right] I \tag{9.58}$$

If on the other hand, the coil is overpitched by an angle β', the fictitious air-gap current components come out to be:

$$I_1^{\mathrm{O}} = \left[\frac{2\left(p - 1 - \dfrac{\beta' p}{2\pi}\right)}{p}\right] I \quad \text{and} \quad I_2^{\mathrm{O}} = \left[\frac{2\left(1 + \dfrac{\beta' p}{2\pi}\right)}{p}\right] I \tag{9.59}$$

Next, we consider the case, when the core-end-surface is infinitely conducting (or effective zero permeability). In this case, the presence of the air-gap is of no significance, and the image of the system would be the same as that due to the continuous conducting surface, and hence the same as shown in the Figure 9.17.

So far we have discussed the two limiting cases, i.e. an infinitely permeable magnetic surface, and an infinitely electrically conducting surface. When the medium is of finite magnetic permeability, we cannot write the two equations corresponding to Eqs. (9.54) and (9.55) directly. We have to go through the stages shown in Figure 9.18, modifying them for the finite permeability effect, particularly for the image of the fictitious air-gap conductor.

9.2 GRAPHICAL METHOD

As in the case of the electrostatic fields, this method is applicable to the two-dimensional and the axi-symmetric problems, with magnetic linearity. Since the Laplacian magnetostatic fields are similar to the corresponding electrostatic fields, we shall not repeat the details of the method which would be identical with what has been described in Sections 5.2–5.2.4.

So we shall restrict our present discussion to some salient points of difference in plotting a two-dimensional magnetic field in three different regions, i.e. (1) an air-space with $\mu = \mu_0$; (2) a region with $\mu >> \mu_0$ (an idealized ferromagnetic medium with high but constant permeability); and (3) a current-carrying region in which $\mu \simeq \mu_0$ (such as copper, or aluminum, or a similar medium).

First we consider the region (1) the air. Figure 9.22 shows the cross-section of such a region.

Let $\Delta\Omega_m$ be the P.D. between the two adjacent magnetic equipotential surfaces. Hence the magnetic field intensity vector **H** is approximately given by $|\mathbf{H}| = \Delta\Omega_m/\delta l_\Omega$.

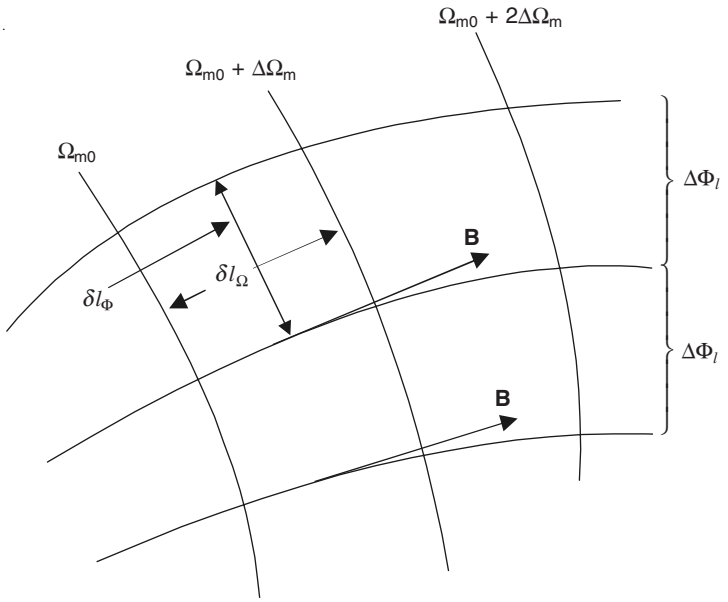


Figure 9.22 A two-dimensional magnetostatic field.

Similarly, from the successive orthogonal lines, let the magnetic flux $\Delta\Phi_l$ (per unit length along the two-dimensional system) be represented by each line of **B**. Then the approximate value of $\Delta\Phi_l$ is given by

$$\Delta\Phi_l = B\delta l_\Phi = \mu_0 \left(\frac{\Delta\Omega_m}{\delta l_\Omega}\right)\delta l_\Phi$$

$$\therefore \quad \frac{\delta l_\Phi}{\delta l_\Omega} = \frac{\Delta\Phi_1}{\mu_0\,\Delta\Omega_m} \tag{9.60}$$

A similar expression would hold for the region (2) of constant permeability in which μ_0 would be replaced by μ. On the interface between the two regions, the **B** vector would refract according to the law given by Eq. (8.45), so that the *B* lines are virtually normal from the air side but not normal on the ferromagnetic side. So 'from the air side only', the surface of the region (2), which is the high permeability region, is nearly an equipotential surface. So such a surface is a very convenient starting point for drawing the curvilinear squares.

Next, in the current-carrying region, i.e. the region (3), the Ω cannot be defined uniquely as explained earlier. But in this medium, the lines of **B** are not refracted at the conductor surface, because there *do not* exist surface currents. Also since the permeabilities of the conductor and the air media are the same, we have both $\mathbf{B}_n$ and $\mathbf{B}_t$ continuous. Hence the mutually orthogonal lines can also be drawn inside the region of the current-carrying conductors. Successive closed lines of **B** will necessarily converge around some point *K*. This point is usually referred to as the 'core' or the 'kernel'. The lines continuing on the lines of magnetic equipotentials, inside the current-carrying conductor are referred to as the 'lines of no work.' So if the position of the kernel is

known, then the plot inside the conductor is also known. But unfortunately, the position of the kernel can be located in advance only in certain special cases. When, however, the position of the kernel has been pre-located, then in the current-carrying region, for successive curvilinear squares,

$$\frac{\delta l'_{\Omega}}{\delta l'_{\Phi}} = \frac{\delta l_{\Omega}}{\delta l_{\Phi}} \left(\frac{S'}{S} \right) \tag{9.61}$$

where ($\delta l'_{\Omega}$, $\delta l'_{\Phi}$) and (δl_{Ω}, δl_{Φ}) are the lengths in the successive curvilinear squares as shown in Figures 9.22 and 9.23, and, S' and S are the areas of the successive 'curvilinear triangles' with

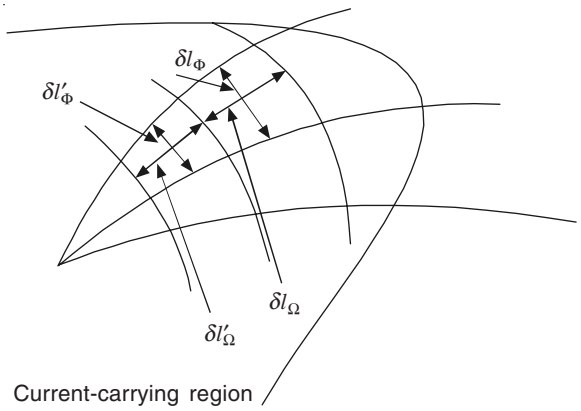


Figure 9.23 Kernel and field-lines within the current-carrying region.

their vertices at the kernel, and the bases as $\delta l'_{\Phi}$ and δl_{Φ} respectively. This expression, along with the trial and error, enables the plotter to locate the kernel. The starting point is always the interface between the conductor and the air-space.

9.3 EXPERIMENTAL METHODS

In practice it becomes necessary to determine the magnetic field in a certain partial region of space only (i.e. the air-gaps or parts of the air-gap of rotating electrical machines). The geometry of the problem might be such that even the approximate methods may not be able to tackle it. Though of course, with the advent of the computer methods and very high speed computers, more and more problems, which were originally thought to be unsolvable, are now being solved. Still the computer solutions often need to be checked by the actual experimental measurements. So, in spite of every thing, the experimental techniques retain their importance.

A very frequently used system, for evaluating the magnetic flux density vector **B** at any discrete set of points, is by what is known as the 'search coil'. It is essentially a very small multi-turn coil wound on a circular, or a square, or a rectangular former (of usually non-magnetic material), such that the coil is quite small compared with the radius of curvature of the field lines. It measures the effects of the magnetic flux, linked by the coil, from which the **B** can be evaluated. However the principle of operation of the device can be understood only on the basis of the phenomenon of the electromagnetic induction, and it will be explained later. We shall now describe another device based on the 'Hall effect'.

9.3.1 Hall Effect Probes

Another method to measure the magnetic flux density at a point is based on what is known as the Hall effect. So we first describe the Hall effect.

The Hall effect was discovered in 1879 at John Hopkins University by E.H. Hall. Since the voltages obtained from the available materials, at that time, were extremely low, no uses outside of the laboratory were possible till suitable materials were developed. The development of high mobility semiconductors has yielded several materials suitable for practical applications of the Hall effect.

The Hall effect is the generation of a voltage across the opposite edges of an electrical conductor carrying a current and placed in a magnetic field. The Lorentz force is the basis of this effect which depends on the deflection of the charged particles moving in a magnetic field. This force is in a direction mutually perpendicular to the path of the particle movement and the direction of the magnetic field.

The basic Hall effect equation is

$$V_H = \left| wR_H \, (\mathbf{J} \times \mathbf{B}) \right|$$

where

V_H = the Hall output voltage

R_H = the Hall coefficient of the material

$\mathbf{J}$ = the current density through the Hall element

$\mathbf{B}$ = the applied magnetic flux density

w = the width of the Hall element.

The vector product ($\mathbf{J} \times \mathbf{B}$) is so written as to indicate the direction sensitivity of the Hall effect. This can also be expressed as

$$V_H = wR_H JB \sin \theta \tag{9.62}$$

where θ is the angle between the direction of $\mathbf{B}$ and the plane of the Hall element as shown in Figure 9.24.

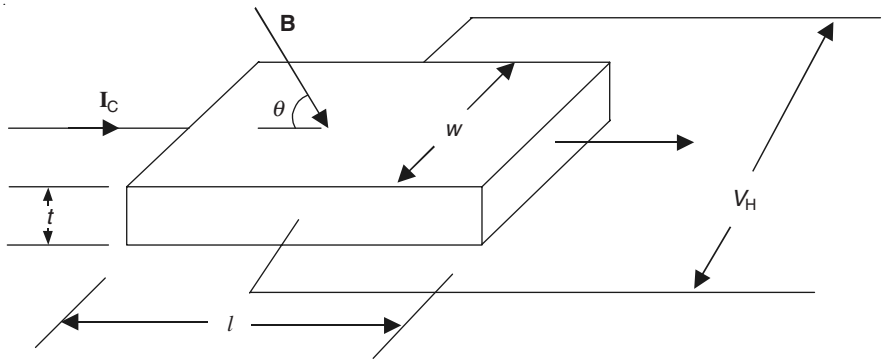


Figure 9.24 The Hall effect.

Since the current density $\mathbf{J}$ is not a convenient parameter, we substitute the total current I_C for it, i.e.

$$J = \frac{I_C}{wt}$$

where I_C is the control current and t is the thickness of the element. Hence the Hall voltage equation becomes

$$V_H = \left(\frac{R_H}{t}\right)\left|\mathbf{I}_c \times \mathbf{B}\right| \tag{9.63}$$

These equations hold for the ideal conditions where the Hall element has an infinite (l/w) ratio. A more useful equation for the practical Hall devices is

$$V_H = K_{HOC}\left|\mathbf{I}_C \times \mathbf{B}\right| \tag{9.64}$$

where K_{HOC} is the open-circuit sensitivity constant taking account of the effects of the geometry and other factors. Strictly speaking K_{HOC} is not entirely independent of $\mathbf{I}_C$ and $\mathbf{B}$ due to the temperature, magneto-resistance, and other effects. But for the present discussion it will be treated as a true constant.

Equation (9.64) shows that the output V_H results from the product of two inputs, and if $\mathbf{I}_C$ is held constant, then V_H is proportional to $\mathbf{B}$. Hence an obvious application of the Hall effect is to measure the magnetic field strength by holding $\mathbf{I}_C$ constant. Many other applications are possible, but at present we note that the device can be used for measuring both the steady and the time-varying magnetic field strength.

The Hall effect is basically a majority carrier mechanism depending on the bulk material properties of the semiconductor material. Unlike transistors, diodes, etc. it is completely independent of the surface effects, junction leakage currents, and junction threshold voltages. Hence the device is highly stable and has a high degree of reproducibility and reliability compared with other semiconductor devices.

From Eqs. (9.62) to (9.64), it is obvious that in order to obtain a high output voltage, the element must have a high Hall coefficient. Also, since the output is proportional to $\mathbf{I}_C$ or $\mathbf{J}$ in the element, its resistance should be as low as practical to prevent excessive heating. In addition, since the noise output is thermal, the low resistance is an important requirement for the device to be used at very low signal levels. Some of the semiconductors used for the Hall elements are indium antimonide (In Sb), indium arsenide (In As) and germanium (Ge). Low resistance, low temperature coefficient, and high output make (In As) the most suitable material for most Hall devices. Hall device elements may also be made of deposited thin films of (In As) and (In Sb). These are mostly used for switching purposes because they do not have the low resistivity property of the bulk materials. The size of the probe is usually quite small, the probe sensing area being usually about 0.02 in^2 and less. The control current is usually 20–200 mA. It is generally advisable to keep the current less than 150 mA to minimize the measurement errors due to self-heating. A typical element can have an open-circuit Hall voltage of 100 mV in the presence of a 10 kG field when a current of 100 mA is applied. The Hall emf is reduced by about 50% when the Hall voltage leads are terminated in a load, such as an indicating meter. The Hall generator can be used to measure permanent magnet fields, alternating fields, and pulsed fields. For alternating and pulsed fields, the current lead arrangement must be made as small as possible to reduce the inductive loop effects. A Hall generator can theoretically operate up to megacycle region. In practice, the Hall element is usually mounted on a substrate of inert material such as alumina (aluminum oxide). The element is generally a few thousandth of an inch thick and needs some sort of mechanical support. The element, the substrate, and the protective encapsulation results in a probe of thickness of the order of 0.020–0.040 inch. The probe is extremely fragile and not repairable.

Some typical probes are shown in Figure 9.25. There are many other applications of the Hall devices, though we have restricted our discussion to magnetic field measurements only.

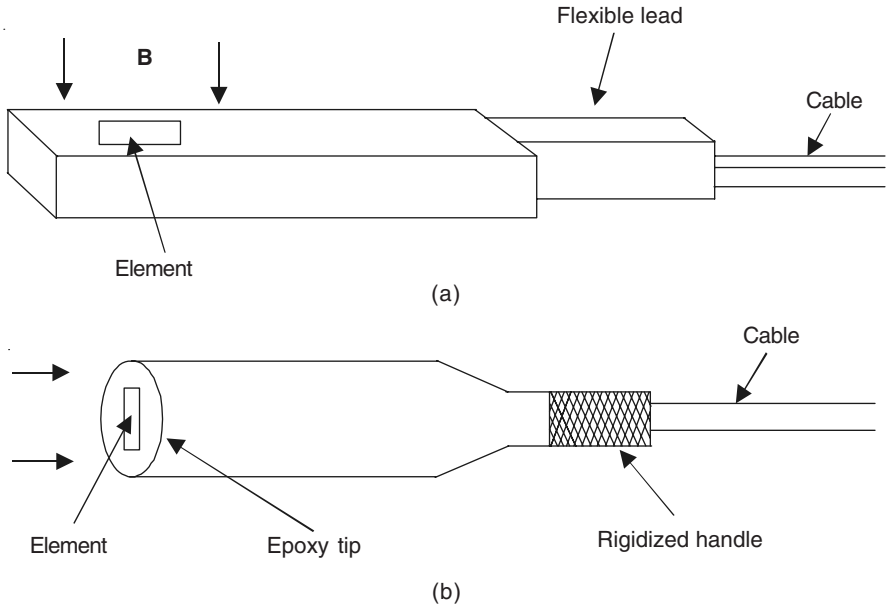


Figure 9.25 Some typical Hall probe configurations: (a) typical flat (transverse) probe and (b) axial probe.

9.3.2 Conducting Paper Analogue

As was mentioned in Section 5.3.2, this method has been very widely used for solving the magnetic field problems. Subsequently, in Section 6.5, the analogy between the electric current and the electric flux was established. Since the magnetostatic fields (not including the current-carrying regions) as well as the electrostatic fields are both Laplacian in nature, the analogy holds for the magnetic fields as well, and hence the current-flow patterns can be used to solve the magnetic field problems.

This conducting sheet method is a very convenient, cheap, and reasonably quick (far quicker than the electrolytic tank method though in no way comparable to the computer methods) method which avoids the tedious trial-and-error process of the graphical field plotting and yet retains the essential simplicity of the method and accelerates it. It does not require much elaborate equipment. Essentially the method consists of establishing in a 'uniformly-conductive' sheet, a field of electrical potential and current flow, which is an analogue representation of the actual field to be investigated. By plotting the equipotentials of the conduction field, or by measuring the potentials at the selected points, information is obtained and is then converted by the appropriate scaling factors to the required parameters of the actual field. Whilst this method can be used to solve different types of field problems, here we shall emphasize the solutions of the electromagnetic fields. Also the method is applicable directly only to the Laplacian fields, though with suitable modifications and some restrictions, it can be used to solve some Poissonian field problems.

9.3.2.1 Analogues for magnetic field studies

The possibility of using a conducting sheet to represent two-dimensional magnetically permeable medium arises from the essential similarity of the equations relating the 'flux-flow' in a magnetically permeable medium, and the current-flow in an electrically conducting medium. That is, the magnetic field equation is

$$\text{M.M.F. (amp-turns)} = \text{Flux} \times \text{Reluctance} \tag{9.65}$$

The electric current flow in the sheet can be expressed in any of the two following ways:

$$\text{Potential (volts)} = \text{Current (amps)} \times \text{Resistance} \tag{9.66}$$

$$\text{Current (amps)} = \text{Potential (volts)} \times \text{Conductance} \tag{9.67}$$

If the problem of Eq. (9.65) is represented by Eq. (9.66), then this representation is a 'direct analogue'; whereas if the problem of the equation is represented by using Eq. (9.67), then an 'orthogonal analogue' is obtained.

9.3.2.2 Mathematical interpretation of the analogues

DIRECT ANALOGY

The equation defining the distribution of the scalar magnetic potential Ω_m in a current-free region is the Laplace's equation

$$\frac{\partial^2 \Omega_m}{\partial x^2} + \frac{\partial^2 \Omega_m}{\partial y^2} + \frac{\partial^2 \Omega_m}{\partial z^2} = 0$$

where Ω_m is the magnetic scalar potential.

The associated equations are

$$\mathbf{H} = -\text{ grad } \Omega_m = -\nabla \Omega_m$$

and

$$\mathbf{B} = \mu \mathbf{H}$$

The equation giving the distribution of the electric potential in a conducting medium has the same form, i.e.

$$\frac{\partial^2 V_e}{\partial x^2} + \frac{\partial^2 V_e}{\partial y^2} + \frac{\partial^2 V_e}{\partial z^2} = 0$$

The associated equations are

$$\mathbf{E} = -\text{ grad } V_e = -\nabla V_e$$

$$\mathbf{J} = \sigma \mathbf{E}$$

There is therefore a direct correspondence between Ω_m and V_e (magnetic and electric potentials), and, **B** and **J** (magnetic flux density and electric surface current density), and a 'direct electrical analogue' can be set up.

If the magnetic and the electric potentials are related by a scale factor 'a', so that

$V_e = a\Omega_m$, then

$$\mathbf{J} = \sigma\mathbf{E} = -a\sigma\nabla\Omega_m = a\sigma\mathbf{H} = \left(\frac{a\sigma}{\mu}\right)\mathbf{B}$$

$$\therefore \qquad \mathbf{B} = \left(\frac{\mu}{a\sigma}\right)\mathbf{J} \tag{9.68}$$

A rigorous detailed scaling analysis for a similar electrostatic system has been described in Section 6.5.

ORTHOGONAL ANALOGY

By comparing Eqs. (9.65) and (9.67), the following orthogonal analogy is obtained:

Magnetic quantity	*Electrical quantity*
1. M.M.F. (amp-turns)	Current (amps)
2. Flux	Potential (volts)
3. Reluctance	Conductance

From the above comparison, we see, from 2, that the electrical equipotentials correspond to the magnetic flux lines, and that 3 indicates that the low reluctance (i.e. high permeability) materials in the magnetic system will be represented by the high resistivity (i.e. low conductivity) materials in the electrical analogue. In particular, the magnetic equipotentials will be represented by the insulating surfaces. Conversely, the surfaces not admitting the magnetic flux (as a consequence of the ideal eddy current screening—a topic to be discussed and explained later in Chapter 15) will be represented by the electrical equipotentials. Since the flux lines of the magnetic system correspond to the equipotentials of the electrical system (i.e. an orthogonal relationship), the flux lines can be plotted directly. So if a 'conjugate electrode' electrical analogue is set up, with the equipotential boundaries of the direct analogue replaced by the insulating boundaries, and the insulating boundaries of the direct analogue replaced by the equipotential boundaries, then the equipotentials of the second analogue will follow the current flow lines of the direct analogue and hence the flux flow lines of the original magnetic system. Similarly, the current flow lines of the second analogue will follow the equipotentials of the direct analogue. Let us now consider the two-dimensional magnetic field and its electrical analogue shown below in Figure 9.26.

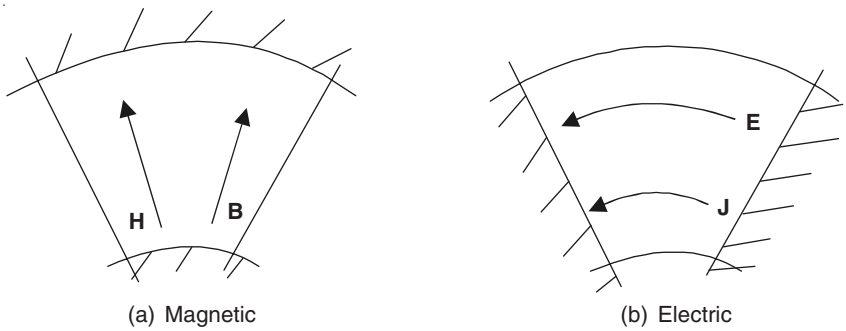


Figure 9.26 Magnetic field problem and its electric analogue (/ / / indicates equipotentials).

The magnetic field intensity in the original problem is given by

$$\mathbf{B} = \mu\mathbf{H}$$

The electric field intensity at the corresponding point of the orthogonal electrical analogue is given by

$$\mathbf{E} = \frac{\mathbf{J}}{\sigma} = \rho\mathbf{J}$$

where

$\mathbf{J}$ = the surface current density (on the analogue) per unit width

σ = the surface conductivity

ρ = the surface resistivity = $1/\sigma$.

Now, if $\mathbf{J} = b\mathbf{H}$, where b = the scaling factor, then

$$\mathbf{E} = b\rho\mathbf{H} = \left(\frac{b\rho}{\mu}\right)\mathbf{B} \qquad \text{and} \qquad \mathbf{B} = \left(\frac{\mu}{b\rho}\right)\mathbf{E} \tag{9.69}$$

The orthogonal analogue is generally more convenient of the two, particularly when dealing with the electromagnetic problems. The directness and the simplicity of the flux density and the flux path determinations are much in its favour; also since the analogue is 'current-fed', there is less dependence on the contact resistance.

9.3.2.3 *Practical details of the conducting paper, associated materials, and the method*

One kind of conducting paper, commercially known as 'teledeltos paper' has electro-sensitive matt gray coating on one side, and the other side is black and glossy. The either side can be used for plotting. The surface resistance of the paper is of the order of 2 kΩ per square. Another type of black conducting paper has the surface resistivity of 4 kΩ per square. While a roll of paper is quite homogeneous, it tends to be anisotropic in nature. The ratio of the resistivity of the paper at right angles (y-direction) to the roll to that along the roll (x-direction) varies from 1.05 to 1.20. In such a medium, the potential V does not satisfy the Laplace's equation but an equation of the type given by

$$\rho_y \frac{\partial^2 V}{\partial x^2} + \rho_x \frac{\partial^2 V}{\partial y^2} = 0 \tag{9.70}$$

where ρ_x and ρ_y are the surface resistivities of the paper in the x- and y-directions, respectively. We shall show how by a suitable scaling transformation, such a paper can be made to solve the Laplacian field problems.

9.3.2.4 *Flow of current from a point source into a sheet of material having different conductivities along the two orthogonal axes*

Initially we assume an isotropic (ideal) sheet of resistive material. If a current flows from a point source into an extensive sheet of the ideal paper, then the lines of force and (the lines of flow) are coincident with the radii emanating from that point and the equipotentials are the concentric circles around this point. This is an orthogonal system satisfying the Laplacian field properties and can be represented by the conjugate functions. Such a field can be characterized by the two functions V and U of the coordinates (x, y) or (r, ϕ). The potential function V expresses the scalar potential at any point. The surfaces or the lines V = constant, will be equipotentials. The flux

function or the flow function U = constant, are the lines of force or flow, and intersect the equipotentials orthogonally. Since such a field is Laplacian,

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = 0 \tag{9.70a}$$

this can be expressed as conjugate functions

$$V - jU = f(x + jy)$$

where

$$\frac{\partial V}{\partial x} - j\frac{\partial U}{\partial x} = f'(x + jy) \quad \text{and} \quad \frac{\partial V}{\partial y} - j\frac{\partial U}{\partial y} = j f'(x + jy)$$

f' being the derivative of f with respect to its argument.

$\therefore$ From the above,

$$\frac{\partial V}{\partial y} = \frac{\partial U}{\partial x} \quad \text{and} \quad \frac{\partial V}{\partial x} = -\frac{\partial U}{\partial y} \tag{9.71}$$

Since the flow function can be expressed in terms of the surface current I, $U = \rho I$ (ρ = surface resistivity of the paper),

$\therefore$
$$\frac{\partial V}{\partial y} = \rho\frac{\partial I}{\partial x} \quad \text{and} \quad \frac{\partial V}{\partial x} = -\rho\frac{\partial I}{\partial y} \tag{9.72}$$

$\therefore$ I also satisfies the Laplace's equation.

$\therefore$ The transformation required for this problem is

$$W = V - jU = V - j\rho I = A \ln (x + jy) \tag{9.73}$$

Or expressing in the cylindrical coordinates, $(x + jy) = re^{j\phi}$

$$V - j\rho I = A \ln r + jA\phi \tag{9.74}$$

$\therefore$ $V = A \ln r$, and the equipotentials are given by r = constant, i.e. concentric circles.

Also, $I = A\phi/\rho$, and the stream lines are the lines radiating from the origin. If the total current entering the sheet (at the origin) is I, then $A = \rho I/(2\pi)$.

$\therefore$ The potential at any point $= -\left(\frac{\rho I}{2\pi}\right) \ln r$

and the potential gradient is

$$-\frac{\partial V}{\partial r} = \frac{\rho I}{2\pi r}$$

Now, if we consider the conducting sheet which has different conductivities (or resistivities) in two orthogonal directions, i.e. ρ_x along the x-axis and ρ_y along the y-axis, then $U \neq \rho I$ and Eq. (9.72) gets modified to

$$\frac{\partial V}{\partial x} = -\rho_x \frac{\partial I}{\partial y} \quad \text{and} \quad \frac{\partial V}{\partial y} = \rho_y \frac{\partial I}{\partial x} \tag{9.75}$$

So the potential gradients along the x- and the y-axes will be proportional to the resistivities in

the respective directions, and the stream lines are no longer coincident with the lines of force, except along the axes. Hence from Eqs. (9.75), we get

$$\rho_y \left(\frac{\partial^2 V}{\partial x^2} \right) = -\rho_y \rho_x \left(\frac{\partial^2 I}{\partial y \partial x} \right) \quad \text{and} \quad \rho_x \left(\frac{\partial^2 V}{\partial y^2} \right) = \rho_x \rho_y \left(\frac{\partial^2 I}{\partial x \partial y} \right)$$

Adding these two equations,

$$\rho_y \left(\frac{\partial^2 V}{\partial x^2} \right) + \rho_x \left(\frac{\partial^2 V}{\partial y^2} \right) = 0 \tag{9.76}$$

If $\rho_x = \rho_y$, then Eq. (9.76) degenerates to the Laplace's equation.

A general solution of Eq. (9.76) is

$$V = F_1 \left[x + j \sqrt{\frac{\rho_y}{\rho_x}}\, y \right] + F_2 \left[x - j \sqrt{\frac{\rho_y}{\rho_x}}\, y \right]$$

Following the analysis of the isotropic medium, the functions V and U (the quasi-conjugate functions now) can be expressed as

$$V - jU = f \left[x + j \sqrt{\frac{\rho_y}{\rho_x}}\, y \right] \tag{9.77}$$

Differentiating Eq. (9.77) with respect to x and y, respectively,

$$\frac{\partial V}{\partial x} - j \frac{\partial U}{\partial x} = f' \left[x + j \sqrt{\frac{\rho_y}{\rho_x}}\, y \right]$$

$$\frac{\partial V}{\partial y} - j \frac{\partial U}{\partial y} = \sqrt{\frac{\rho_y}{\rho_x}}\, f' \left[x + j \sqrt{\frac{\rho_y}{\rho_x}}\, y \right]$$

Equating the real and the imaginary terms,

$$\frac{\partial V}{\partial x} = -\sqrt{\frac{\rho_x}{\rho_y}} \left(\frac{\partial U}{\partial y} \right)$$

and (9.78)

$$\frac{\partial V}{\partial y} = -\sqrt{\frac{\rho_y}{\rho_x}} \left(\frac{\partial U}{\partial x} \right)$$

These equations are identical with Eqs. (9.75).

$$\therefore \quad U = I \sqrt{\rho_x \rho_y} \tag{9.79}$$

Hence Eq. (9.77) becomes

$$V - jI\sqrt{\rho_x \rho_y} = f \left[x + j \sqrt{\frac{\rho_y}{\rho_x}}\, y \right] \tag{9.80}$$

Thus, the corresponding logarithmic transformation for the flow in this anisotropic medium is

$$V - jI\sqrt{\rho_x \rho_y} = A \ln\left[x + j\sqrt{\frac{\rho_y}{\rho_x}}\, y\right] \tag{9.81}$$

which corresponds to Eq. (9.73) for the isotropic medium.

Writing

$$x + j\sqrt{\frac{\rho_y}{\rho_x}}\, y = a(\cos\alpha + j\sin\alpha) = ae^{j\alpha}$$

where a and α will be related to the polar coordinates r and ϕ as shown below in Figure 9.27 (assuming $\rho_y > \rho_x$), i.e.

$$a = \sqrt{x^2 + \frac{\rho_y}{\rho_x} y^2} \qquad \text{and} \qquad \tan\alpha = \frac{y}{x}\sqrt{\frac{\rho_y}{\rho_x}}$$

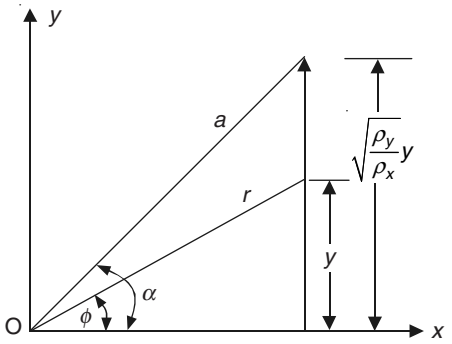


Figure 9.27 Relationship between (r, ϕ) and (a, α).

Then

$$V - j\sqrt{(\rho_x \rho_y)}I = A \ln a + jA\alpha$$

and

$$V = A \ln a, \qquad I = -\frac{A\alpha}{\sqrt{\rho_x \rho_y}}$$

The equipotentials are then given by ln a = constant, or a = constant, i.e.

$$x^2 + \left(\frac{\rho_y}{\rho_x}\right) y^2 = a^2 \tag{9.82}$$

Thus, if a is a constant, Eq. (9.82) is an ellipse of semi-major axis a and semi-minor axis $a\sqrt{\rho_x/\rho_y}$. All the ellipses have the same ratio of the axes (i.e. confocal ellipses). The stream lines are given by α = constant, and these lines radiate from the origin.

The stream lines and the equipotentials are shown in Figure 9.28. The lines of force are of course normal to the equipotentials, but the lines of flow are no longer normal to the equipotentials. This is because when $\rho_y > \rho_x$, the material conducts better parallel to the axis of x than it does parallel to the axis of y. At the point P, shown in Figure 9.28, the electric force is

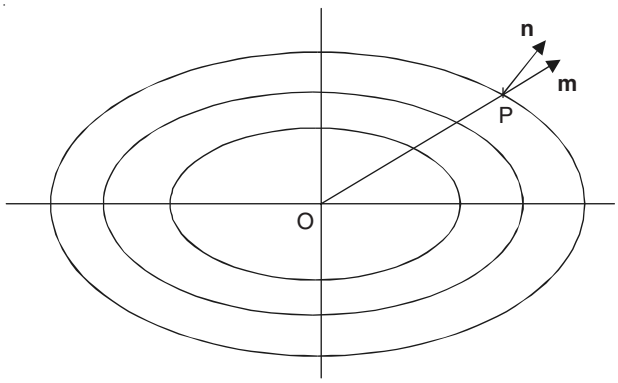


Figure 9.28 Equipotential ellipses.

along **n** (normal to the ellipse). But because of the better conductivity in the x-direction, the resulting current flows in the direction **m**.

If the total current flowing in at the source O is I, then

$$A = -\left(\frac{\sqrt{\rho_x \rho_y}}{2\pi}\right) I$$

and the potential V at any point is

$$V = -\left(\frac{\sqrt{\rho_x \rho_y}}{2\pi}\right) I \ln a$$

$$= -\left(\frac{\sqrt{\rho_x \rho_y}}{2\pi}\right) I \ln (r \cos \phi \sec \alpha)$$

$$= -\left(\frac{\sqrt{\rho_x \rho_y}}{2\pi}\right) I [\ln r + \ln (\cos \phi \sec \alpha)]$$

$\therefore$ The potential gradient along the radius vector is

$$\frac{\partial V}{\partial r} = -\left(\frac{\sqrt{\rho_x \rho_y}}{2\pi r}\right) I \tag{9.83}$$

This is the same as in an isotropic sheet, from which, it follows from the above equation that the 'potential gradient along the radius vector is independent of its orientation'. This conclusion is somewhat unexpected, but has been confirmed experimentally. To explain this point, we consider four electrodes A, P, Q, B placed in a line on an extensive plate having different conductivities in two different orthogonal directions, as shown in Figure 9.29.

A given current I is fed in at the point A, and taken out at B. The potential difference between the points P and Q is independent of the orientation of the line on the plate. This is because, if the direction AB is highly resistive, then the current spreads out very much in the direction at right angles to AB, and the current density along PQ is quite small. Thus V_{PQ} is due to a low current density through a material of high resistivity. If, now, APQB is turned through a

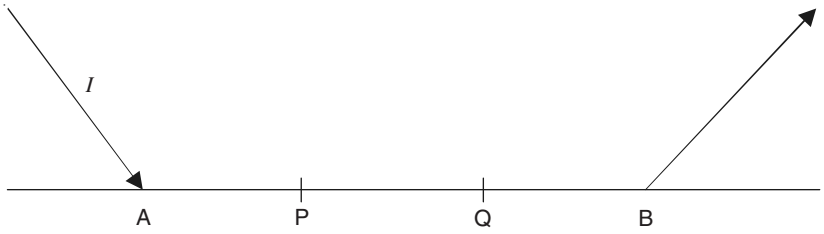


Figure 9.29 Current through a line in an anisotropic medium.

right angle, the same total current gives a higher current density between P and Q, which with lower resistivity gives the same P.D. as before.

Thus, we have seen how the effect of the anisotropic nature of the conducting paper can be corrected for the representation of the Laplacian field problems. But more than this point, it opens a new avenue for analyzing the fields in the anisotropic media. This is an important practical aspect, because in the present-day large generator-transformers, the cores are made up of CRGO (cold rolled grain oriented) steel laminations which are highly anisotropic in nature. The laminations are so cut from the roll that both in the vertical limbs as well as in the top and the bottom horizontal sections of the transformer cores, the magnetic flux flows along the preferred low reluctance direction of the material. In these regions, the magnetic flux does not obey the Laplacian equation, and instead satisfies Eq. (9.70). Using the scaling factor based on the present analysis, the problem can now be reduced to that of a Laplacian field which can then be directly analyzed.

9.3.2.5 Method of obtaining the field plot

Thus Eq. (9.69), which we rewrite below,

$$\rho_y \frac{\partial^2 V}{\partial x^2} + \rho_x \frac{\partial^2 V}{\partial y^2} = 0$$

can now be reduced to the Laplace's equation by the substitution

$$x_1 = x \qquad \text{and} \qquad y_1 = y \sqrt{\frac{\rho_y}{\rho_x}} \tag{9.84}$$

Thus to account for the anisotropy, whether of the paper or of the material characteristic, it is necessary to compress the scale at right angles to the roll direction by the square root of the resistivity ratio. The resistivity ratio for a given roll of paper (or material) can easily be determined by the potential measurements on an L-shaped sample cut with the two limbs of the L along and across the roll direction. The paper should be ideally laid on an insulating base.

A suitable scale (or the scales for the anisotropy) having been chosen, the problem is drawn on the conducting paper. Note that india rubber should not be used to rub out incorrect lines. The paper is then cut along the marked outline. The equipotential boundaries and the current sources are painted with a silver conducting paint. The paint must be thoroughly mixed, and a few coats of the paint must be given on the equipotential boundaries. After a couple of hours of drying, the P.D. along the equipotential has to be checked. In order to ensure the equipotentiality of these regions, it might be preferable to fix a stripped thin wire along such lines, and then paint on the wire. Once the preparation is complete, the field-plotting potentiometer can be connected across

the painted electrodes of the model, and the equipotentials plotted at the intervals of equal P.D. If it is a direct analogue, the equipotentials are the equipotentials of the original problem, and the flux lines have to be plotted by the method of curvilinear squares. On the other hand, if it is an orthogonal analogue, then we have got the flux lines directly.

This method can be directly extrapolated to represent and solve multi-electrode analogues as well.

9.3.2.6 Representation of Poissonian fields

The method we have discussed so far, deals with only Laplacian fields. There are three main ways of extending this technique to solve the Poissonian field problems. We shall briefly describe their underlying principles.

CAPACITIVELY-COUPLED POISSONIAN FIELD PLOTTING

Because this method is based on the concept of the 'displacement current,' the explanation given below will be understood properly only after the study of the Maxwell's equations in Chapter 12. So the reader can proceed to the next section and study this explanation later. This method utilizes a resistive plotting surface, and a coupling capacitor. The coupling capacitor is formed by the resistive plotting surface and a parallel coupling electrode. A displacement current acting through this coupling capacitor sends a distributed current into the resistive surface. The voltage distribution created by this means (i.e. due to an alternating current of sufficiently low frequency, capacitively-coupled to the paper surface) will satisfy a Poissonian equation involving the resistance per square of the paper (= K) and the constant displacement current density J_C for the defined source region. For each capacitively-coupled region, the Poissonian equation $\nabla^2 V(x, y) = KJ_C$ applies, and outside of any such region, the potential distribution satisfies the Laplacian equation $\nabla^2 V(x, y) = 0$. The potential distribution $V(x, y)$ describes the flux distribution as defined by its magnetic vector potential $A_z(x, y)$. Also the resistance of the paper has to be analogous to the μ of the magnetic system. The constant displacement current density J_C can be created in the analogue, if the capacitive reactance of the respective region is much greater than the surface resistance of the paper. J_C can be controlled by the spacing between the electrode and the paper (Figure 9.30). Typical excitation frequencies used are 1–2 kHz. Care has to be taken to glue the teledeltos paper to the glass, for any inhomogeneity of the paper results in large errors in the surface voltage distribution.

MULTIPLE PIN DC ANALOGUE

In this case each source region is a matrix of pins, each delivering a constant current in the conductive sheet. There is a direct correspondence between the magnitude of the effective current density in a source region with that of the real field. In spite of the simplicity of using the direct current, there are some disadvantages in this method. There is a need for an elaborate set-up for accurate results, and hundreds of pins have to be inserted in the conducting paper. The procedure of actually connecting the pins and adjusting each current-feed point is tedious.

POISSONIAN FIELD ANALOGUE BY USING A MECHANICAL INTEGRATOR

Most of the disadvantages of the previous two methods can be eliminated by using this method. The basis of this method depends on the fact that the solution of the Poisson's equation consists of two parts—the complementary function (C.F.) and the particular integral (P.I.), where the

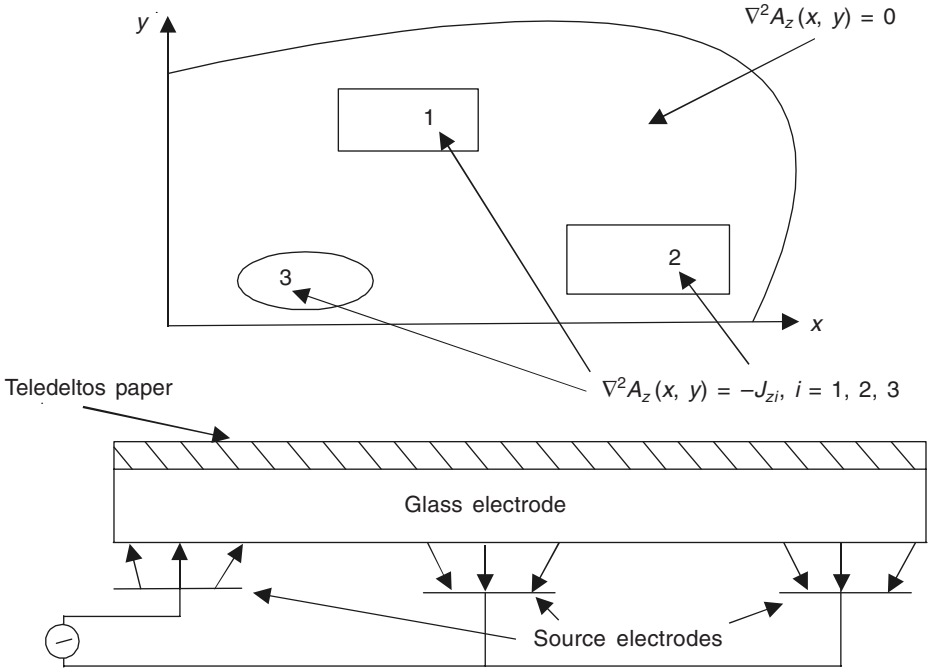


Figure 9.30 A multiple-source Poissonian and Laplacian field problem and its capacitive analogue.

(C.F.) is the solution of the corresponding Laplace's equation. The two-dimensional Poisson's equation is

$$\nabla^2\Omega = \frac{\partial^2\Omega}{\partial x^2} + \frac{\partial^2\Omega}{\partial y^2} = K$$

where K is a known constant of $f(x, y)$.

For the C.F., the right-hand side (of the Poisson's equation) = 0 , and let the C.F. solution be ϕ, then:

$$\frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} = 0 \qquad \text{(Laplace's equation)}$$

$\therefore$ $$\Omega = \text{C.F.} + \text{P.I.}$$

$$\nabla^2\Omega = \nabla^2\phi + \nabla^2(\text{P.I.}) = \nabla^2(\text{P.I.}) = K$$

$\therefore$ $$\text{P.I.} = \left(\frac{1}{\nabla^2}\right)K = \left(\frac{1}{D_x^2 + D_y^2}\right)K = \left(\frac{K}{2}\right)(x^2 + y^2)$$

assuming K to be constant; and where D_x, D_y are the partial derivatives with respect to x and y respectively.

$\therefore$ The general solution of the Poisson's equation is

$$\Omega = \phi + \left(\frac{K}{2}\right)(x^2 + y^2)$$

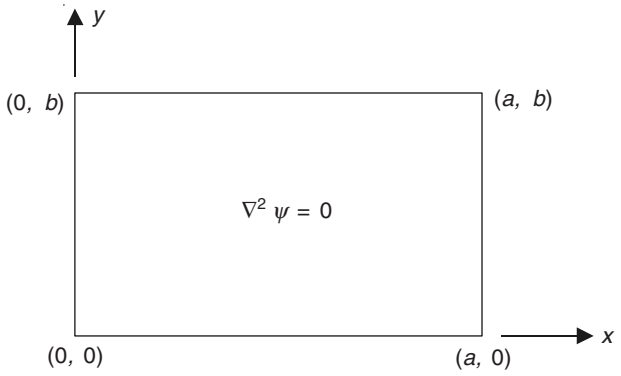


Figure 9.31 A Poissonian field reduced to Laplacian field.

∴ The required substitution to convert the Poisson's equation to the Laplace's equation is

$$\Psi = \Omega - \left(\frac{K}{2}\right)(x^2 + y^2)$$

Then the Poisson's equation reduces to $\nabla^2 \Psi = 0$, which is the Laplace's equation.

Let the region under consideration be a rectangular region as shown in Figure 9.31.

We shall now show how the boundary conditions have to be modified.

(a) *Dirichlet-type boundary.* Let, on $x = 0$, $\Omega = 0$.

This changes to

$$\Psi = \Omega - \left(\frac{K}{2}\right)(x^2 + y^2) = -\left(\frac{K}{2}\right) y^2$$

(b) *Normal flux.* Let on $x = 0$, only normal flux exist; i.e.

$$B_x = \frac{\partial \psi}{\partial x} \text{ exists and so } B_y = \frac{\partial \Omega}{\partial y} = 0$$

$$\therefore \quad \frac{\partial \psi}{\partial y} = \frac{\partial \Omega}{\partial y} - \left(\frac{K}{2}\right)(0 + 2y)$$

$$\therefore \quad \text{On } x = 0, \ \frac{\partial \psi}{\partial y} = -Ky$$

(c) *Tangential flux.* Let on $x = 0$, only tangential flux exist; i.e.

$$B_y = \frac{\partial \Omega}{\partial y} \text{ exists and so } B_x = \frac{\partial \Omega}{\partial x} = 0$$

$$\therefore \quad \text{On } x = 0, \ \frac{\partial \psi}{\partial x} = -Kx$$

The Laplacian field can be set up on the conducting sheet (the teledeltos paper) as described before. To obtain the Poissonian field, all we need to do is to modify the boundaries to meet the new conditions as derived just now. For this purpose, a mechanical integrator is used,

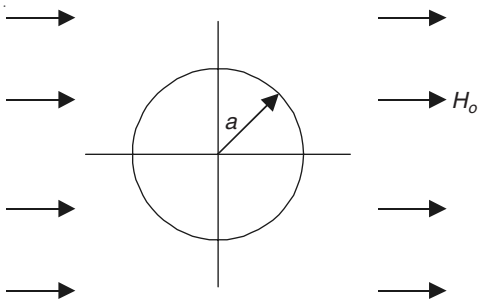
which is a stainless steel strip of ideally infinite length (for practical purposes, of sufficient length, and the length is >> the width). By applying a voltage of V volts across the ends of the length of the strip, the potential drops from V volts at one end to 0 at the other. By taking the current leads from along the length of the strip at suitable intervals, and feeding them into the boundary of the conducting sheet at the points so scaled as to meet any of the conditions derived, we can produce the requisite boundary condition on the specified boundary. The current at each point can be controlled for the required function, by feeding into the model through suitable resistors. The currents have to be controlled carefully to maintain the required accuracy of the model.

9.4 APPROXIMATE METHODS

The approximate numerical methods described for the electrostatic field problems—the finite difference methods and the finite element method—are equally applicable to the problems of the magnetostatic fields. For the Laplacian and the Poissonian fields, the methods are identical, and we shall not repeat the details here again. The main difference between the electrostatic and the magnetostatic field problems is that whilst for the electrostatic fields, the source is a scalar quantity, the source for the magnetic fields is a vector quantity. Though, of course, it is possible to use the concept of the scalar magnetic potential for the magnetic fields, it does introduce certain constraints and also limits the range of the problems that can be solved. Since so far we have not studied the concept of the 'vector potential' in detail and in a rigorous manner, we shall postpone the discussion of the methods applied to the 'vector potential' for a later stage. Furthermore, when we have also studied the time-varying fields (i.e. the electromagnetic fields), it will be seen that the magnetostatic field problems can be considered as special and restricted cases of the general field problems. For example, using the finite element method, the general 'functional' for the electromagnetic fields can always be simplified (by dropping certain terms) to obtain the solutions of the magnetostatic problems. We end our present introductory comments at this stage, by reminding ourselves that while using both the F.E. method as well as the F.D. method, we should be careful during the conversion of the magnetic boundary conditions to their mathematical equivalents, i.e. to make the correct correspondence between the physical conditions and their mathematical interpretations.

PROBLEMS

9.1 A solid cylinder of radius a and relative permeability μ_r as shown in the adjoining figure is placed in a homogeneous magnetic field H_o whose direction is perpendicular to the axis of the cylinder. Assume the cylinder to be infinitely long axially and hence neglect the edge effects. Obtain the expressions for the magnetic scalar potential Ω both inside and outside the cylinder, and show that the magnetic intensity inside the cylinder is



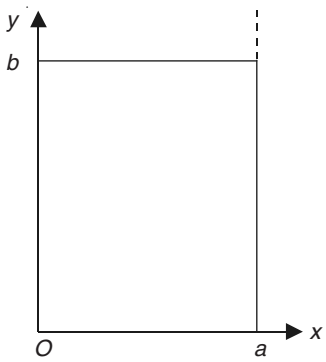
$$H_{(r < a,\ \phi)} = \frac{2H_o}{1 + \mu_r}$$

Hint: In two-dimensional cylindrical polar coordinate system, with the origin at the centre of the cylinder, Laplace's equation in r and ϕ variables needs to be solved. The original undisturbed field H_o can be shown to be expressed by the scalar potential Ω in the form

$$\Omega_o = H_o r \cos \phi$$

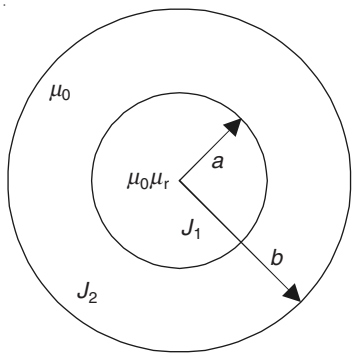
Solve the problem for both Ω_{in} and Ω_o for the two regions.

9.2 A rectangular strip of metal as shown in the adjoining figure is held at zero potential at the bottom edge and has a potential distribution $\Omega = V$ at the top edge. The other two sides are insulated, so that there is no normal component of current on these boundaries. State the boundary conditions in mathematical terms. Evaluate the potential distribution in the strip.



Hint: In which coordinate direction, the potential distribution would consist of orthogonal functions? Why?

9.3 The adjoining figure shows the cross-section of a composite conductor used in overhead transmission lines. It consists of a steel inner wire of radius a, and an annular outer conductor of radius b, both being in electrical contact. Calculate the magnetic field within the conductors.

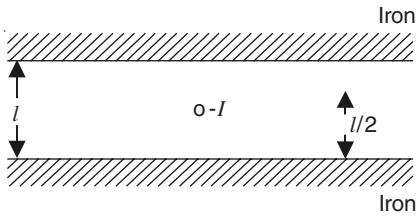


Hint: For simplicity, assume the current density to be uniform in both the conductors, but different (their conductivities being σ_1 and σ_2 respectively. Write down the expression for the total current I in terms of the two current densities J_1 and J_2.

This problem can be solved either by solving one-dimensional Laplace's equation in r, or by Ampere's law which is simpler. However solve by both the methods and ensure the correctness of the solution. State the interface boundary (or continuity) conditions.

9.4 Show that the magnetic potential due to a linear current I situated midway in the airspace bounded by two plane parallel walls of infinitely permeable iron is

$$\Omega = 2I \tan^{-1}\left[\frac{\tan\left(\pi y/l\right)}{\tanh\left(\pi x/l\right)}\right]$$



Hint: Use the standard integral $\int \frac{d\theta}{\alpha + \beta \cos\theta}$.

10 Time-Varying Fields and Electromagnetic Induction

10.1 THE TOTAL FORCE BETWEEN TWO SMALL MOVING CHARGES

In Chapter 7, Sections 7.1 to 7.3, we have seen that the total force between two small 'moving' charges cannot be given by the Coulomb's law alone. Certain corrective terms had to be added to account for the additional force. It was also noted that for the reasonable velocities of the charges, the corrective terms are very small compared with the Coulomb's force and that their experimental determination was not possible. So these additional terms were determined in an indirect manner. What we did was to measure the total force on a large number of charges in a known state of 'steady' motion, and then reduced mathematically to the contributions of pairs of individual moving charges. This additional term was

$$\mathbf{F}_{m12} = \mu_0 \left(\frac{Q_1 Q_2}{4\pi} \right) \left[\frac{\mathbf{v}_2 \times (\mathbf{v}_1 \times \mathbf{u}_{12})}{r^2} \right] \tag{10.1}$$

and this was referred to as the 'magnetic force.' The terms $\mathbf{v}_1$ and $\mathbf{v}_2$ represent the velocities of the charges Q_1 and Q_2 respectively, and r represents the distance between the charges, and $\mathbf{u}_{12}$ the unit vector along r in the direction from Q_1 to Q_2, and $\mathbf{F}_{m12}$, the additional force by which the moving charge Q_1 acts on the moving charge Q_2. But this is ***not*** the total additional force between the two charges in arbitrary state of motion, not even when the two isolated charges are moving with uniform velocities. So in addition to the steady-state additional force term, another corrective term is required. For this purpose, to obtain the total force between the two charges (in slow motion), the forces between the non-steady current systems have to be investigated. This is done by the experiments with the non-steady motion of a large number of charges, i.e. by considering the mutual effects of the non-steady current systems in which both the geometry and the current density vary with time. The simplest experiments are those with the varying currents and charges in stationary conductors.

Later we shall see when the currents and the charges are varying with time, the fields, produced by them, do not change in synchronism with them, and their changes are always retarded (i.e. lagging behind) with respect to the changes in the currents and the charges.

However, if the sources (i.e. the currents and the charges) vary slowly with the time, and the distances concerned are not very large, then the retardation effects in their interactions can be neglected. At this stage, we are concerned with the time-variations in which the retardation effects can be justifiably ignored, i.e. the changes in the currents and the charges produce changes in their fields practically simultaneously. Such systems are also referred to as the 'quasi-stationary systems'. We shall have to consider the retardation effects at higher velocities, later, i.e. microwaves and antennae.

10.1.1 The Force between the Two (Slow) Small Charges in an Arbitrary State of Motion

The correcting term for the forces acting between all the systems of the (quasi-stationary) macroscopic electric currents and the charges is of the form

$$\mathbf{F}'_{e12} = -\mu_0 \left(\frac{Q_1 Q_2}{4\pi}\right) \frac{\partial}{\partial t}\left(\frac{\mathbf{v}_1}{r}\right)_{\mathbf{v}_2 = 0} \tag{10.2}$$

where $\mathbf{F}'_{e12}$ is the force exerted by a charge Q_1 on a charge Q_2. The charge Q_2 is considered to be stationary at the instant of consideration, which is why the subscript $\mathbf{v}_2 = 0$ is used for the time-derivative term. The force term $\mathbf{F}'_{e21}$ exerted by Q_2 on Q_1 is obtained by interchanging the subscripts 1 and 2.

Hence from the macroscopic experiments, with the (slowly) varying currents and the clusters of charges, the total force exerted by a charge Q_1 on a charge Q_2 is

$$\mathbf{F}_{12} = \left(\frac{1}{4\pi\varepsilon_0}\right)\left(\frac{Q_1 Q_2}{r^2}\right)\mathbf{u}_{12} + \left(\frac{\mu_0}{4\pi}\right)\left(\frac{Q_1 Q_2}{r^2}\right)\left[\mathbf{v}_2 \times (\mathbf{v}_1 \times \mathbf{u}_{12})\right] - \left(\frac{\mu_0}{4\pi}\right)(Q_1 Q_2)\frac{\partial}{\partial t}\left(\frac{\mathbf{v}_1}{r}\right)_{\mathbf{v}_2 = 0} \tag{10.3}$$

The force $\mathbf{F}_{21}$ exerted by the charge Q_2 on the charge Q_1 is obtained from this equation by interchanging the subscripts 1 and 2. It will be seen that in general $\mathbf{F}_{12} \neq \mathbf{F}_{21}$. The reason for this is that we cannot consider the moving two point charges as a two-body problem, and the system has to be considered as a three-body problem.

10.1.2 The Physical Meaning of the Electromagnetic Field

The expression for the total force $\mathbf{F}_{12}$, given by Eq. (10.3), can be split into two parts—one part depending explicitly on the velocity $\mathbf{v}_2$ of the charge Q_2 and the other independent of $\mathbf{v}_2$. Thus,

$$\mathbf{F}_{12} = Q_2\left[\left(\frac{1}{4\pi\varepsilon_0}\right)\left(\frac{Q_1}{r^2}\right)\mathbf{u}_{12} - \left(\frac{\mu_0 Q_1}{4\pi}\right)\frac{\partial}{\partial t}\left(\frac{\mathbf{v}_1}{r}\right)_{\mathbf{v}_2 = 0}\right] + Q_2\mathbf{v}_2 \times \left[\left(\frac{\mu_0}{4\pi}\right)\left(\frac{Q_1}{r^2}\right)(\mathbf{v}_1 \times \mathbf{u}_{12})\right] \tag{10.4}$$

The first term in the first bracket is the coulomb field intensity produced by the charge Q_1, the only difference between this term and the electrostatic force expression is that this quantity varies with time as r changes with time. The second term in the first bracket is an additional electric field intensity created by a 'moving charge'. So we can now define the 'total electric field intensity' of a point charge (with respect to an observer) as

$$\mathbf{E} = \left(\frac{1}{4\pi\varepsilon_0}\right)\left(\frac{Q}{r^2}\right)\mathbf{u} - \left(\frac{\mu_0 Q}{4\pi}\right)\frac{\partial}{\partial t}\left(\frac{\mathbf{v}}{r}\right) \tag{10.5}$$

v being the instantaneous velocity of the charge, and the unit vector **u** is directed conventionally *away* from the charge.

The third term of Eq. (10.4) is the already known expression for the force on a charge Q_2 moving in the magnetic field of the charge Q_1. The magnetic flux density vector **B** created by a point charge Q, moving with the instantaneous velocity **v**, is given by

$$\mathbf{B} = \left(\frac{\mu_0}{4\pi}\right)\frac{Q(\mathbf{v}\times\mathbf{u})}{r^2} \tag{10.6}$$

where **u**, **v**, and r have the same meanings as in Eq. (10.5). Thus the total force exerted on Q_2 by a charge Q can be written in the same form as the Lorentz force, i.e.

$$\mathbf{F}_{\text{on } Q_2} = Q_2\mathbf{E} + Q_2\mathbf{v}_2 \times \mathbf{B} \tag{10.7}$$

where **E** stands for the generalized electric field intensity defined by Eq. (10.5), and subsequently we shall always use **E** in this sense. For the static charges, it is identical with the electric field intensity defined before. In the generalized situation, **E** is not necessarily a vector function of the space coordinates only, and is not necessarily conservative.

These considerations enable us to draw an important additional conclusion. The force on Q_2 given by Eq. (10.3) or (10.7) is a 'physical entity'. It has been split into two distinct parts, which are derived from the two vector functions of the coordinates and of the time, and named as the electric and the magnetic field (as a matter of convenience only). But what we must understand is that *the electric and the magnetic fields are two parts of one unique physical phenomenon.* There is a simultaneous existence of both the components in all cases, though one of them may almost completely be suppressed in some regions of space. For example, in an electrostatic field, we say that there is no magnetic field which is basically not quite correct. This is because at the sub-atomic level, the electrons are in a perpetual Brownian-like motion. Thus, even in this case, the magnetic field does exist, though macroscopically it is negligible.

Similarly, in a permanent magnet, a microscopic electric field does exist in the vicinity of each atom, though again at the macroscopic level, it (the electric field) is negligible. Thus the unique physical field, which has been split as a matter of convenience, into two parts, and referred to as the generalized electric and magnetic fields, is known as the 'electromagnetic field'. The two electrostatic and magnetostatic fields, which we have been talking about so far, are in fact, the two extreme or the limiting cases of this electromagnetic field.

10.2 ELECTROMAGNETIC INDUCTION

We now consider a (quasi-stationary) changing or varying current system. We have stated earlier that in such a system, the changes in the source and the consequent changes in the resulting fields can be considered nearly simultaneous. We shall see later that the systems can be considered quasi-stationary up to frequencies in the MHz range.

Such systems include the changes produced by the slowly moving, current-carrying conductors, or the stationary conductors with the slowly varying currents, or a combination of both. Initially, we shall consider the stationary conductors or closed loops.

In such a system, the current density **J** is a function of time. Let Q be the charge of the free charge carriers, **v** their average velocity, and N the number of free charge carriers per unit volume. Then, by Eq. (10.5) and the principle of superposition, the electric field intensity created by these moving charges is

$$\mathbf{E} = -\left(\frac{\mu_0}{4\pi}\right)\sum_v \left[Q\frac{\partial}{\partial t}\left(\frac{\mathbf{v}}{r}\right)\right] + \text{a coulomb force term} \tag{10.8}$$

Macroscopically this sum can be represented as an integral, i.e.

$$\mathbf{E} = -\left(\frac{\mu_0}{4\pi}\right)\iiint_v \left[\frac{\partial}{\partial t}\left(\frac{NQ\mathbf{v}}{r}\right)dv\right] + \text{a coulomb force term} \tag{10.9}$$

But $NQ\mathbf{v} = \mathbf{J}$, the current density, and the volume v is large enough to enclose all the moving charges, and hence is constant in time. Hence $\mathbf{E}$ can be expressed as

$$\mathbf{E} = -\frac{\partial}{\partial t}\left[\left(\frac{\mu_0}{4\pi}\right)\iiint_v \left(\frac{\mathbf{J}d}{r}\right)\right] + \text{a coulomb force term} \tag{10.10}$$

We shall show later (in Chapter 13), that the (box) bracketed term above is the magnetic vector potential $\mathbf{A}$ (at the instant of time considered). Hence,

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} + \text{a coulomb force term (possible)} \tag{10.11}$$

where $\mathbf{A}$ is related to the magnetic flux density by the equation

$$\mathbf{B} = \text{curl } \mathbf{A} = \nabla \times \mathbf{A}$$

10.2.1 Faraday's Law of Electromagnetic Induction

Induction of currents by a changing magnetic field (*in a stationary closed loop*).

From Eqs. (10.10) and (10.11), it is obvious that when a closed circuit or loop of wire is placed in the changing electric field mentioned in Section 10.2 (i.e. the $\mathbf{E}$ vector in the quasi-static field), this field would act on the free charges inside the loop. The presence of the loop or the closed circuit has the effect of taking a line integral of the field vectors along the contour of the circuit. 'The possible coulomb force term' of Eqs. (10.10) and/or (10.11) will not contribute anything to it—its line integral about the closed loop will be zero [Section 1.6, Eq. (1.28)], but the other term, due to the moving charges, acts as an impressed field. As we shall see later, that the line integral of the magnetic vector potential $\mathbf{A}$ is nothing but the magnetic flux through the contour. The time variation of this appears as an emf at the ends of the loop, i.e.

$$\mathcal{E} = \oint_C \mathbf{E}\cdot d\mathbf{l} = -\oint_C \left(\frac{\partial \mathbf{A}}{\partial t}\right)\cdot d\mathbf{l} = -\frac{\partial}{\partial t}\left(\oint_C \mathbf{A}\cdot d\mathbf{l}\right)$$

$$= -\frac{\partial \Phi}{\partial t} = -\frac{\partial}{\partial t}\left[\iint_S \mathbf{B}\cdot d\mathbf{S}\right] \tag{10.12}$$

We shall see now, how historically this conclusion was arrived at on the basis of some great experimental work done by some of the early pioneers in the field of electrical engineering, and how the science of 'electromagnetism' in the proper sense of the word, was developed. It was in the year 1831, that Michael Faraday made this fundamental and important discovery that 'an

electrical current is induced in a closed circuit, when a magnet in its vicinity is moved.' It was by this discovery that Faraday completed the union of the sciences of current electricity and magnetism, a unification which got started by the experiments carried out by Oersted in 1820, when he, for the first time, observed the magnetic effects associated with an electric current. Subsequently, it was found that the induced current could be produced by more ways than the one mentioned above. Then instead of moving a magnet in the vicinity of a stationary coil, the same effect was also produced when a coil was moved in the neighbourhood of a fixed magnet. Also the same effect was observed in the circuit of a wire loop placed in a magnetic field, and the loop was distorted so as to change its area. Again the induction took place without any movement of the loop or the magnet, if the intensity of the magnetic field was changed by altering the current in a neighbouring coil.

The essence of all such experimental observations is summarized by the following two laws, which were enunciated in about 1845.

1. Neumann's law. "When the magnetic flux linked with a coil (or circuit) is changed in any manner, then an emf is set up in the circuit such that it (the emf) is proportional to the rate of change of the flux-linkage with the circuit".

2. Lenz's law. "The direction of the induced emf is such that any current which it produces, tends to oppose the change of flux".

The Lenz's law is only a particular case of a very general physical principle named **Le Chatelier's Principle** which states that "a physical system always reacts to oppose any change that is imposed from outside." In the present case, the physical change is the change in the magnetic flux linkage and the reaction is the induced emf or the back emf, which opposes the change causing it.

At this stage, we shall make a few comments to explain properly the idea of the flux-linkage. When the circuit is a simple, single-turn one, then the 'flux-linkage' is the amount of the flux Φ which is passing through the single-turn coil. However if it is a multi-turn coil (say, N turns), it is not always necessary that the flux linked is likely to be the same through every turn. The quantity whose rate of change measures the induced emf is in fact the sum of the fluxes through all the turns. This is stated as the 'total flux linkage' with the coil. The notation Φ is used to denote the flux-linkage with a coil as well as through a loop. However in some problems, it is more convenient to use Φ as the flux linked with each turn, and hence the total flux-linkage then becomes ($N\Phi$).

Now, we explain the significance of the Lenz's law with reference to Figure 10.1. In this figure, if the resultant flux through the circuit has the direction A_1 and is increasing, then the

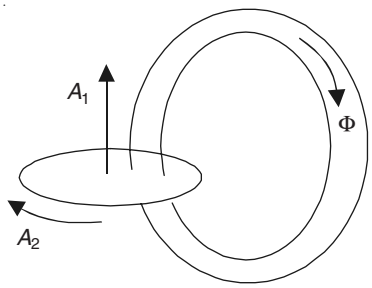


Figure 10.1 Changing flux-linkage and Lenz's law.

induced current in the circuit will tend to set up a flux in the direction opposite to A_1. Hence the direction of the current must be as shown by the arrow A_2.

So, now, we state formally what is known as the Faraday's law of induction:

Faraday's law is initially regarded as giving the emf induced in a wire loop C (Figure 10.2).

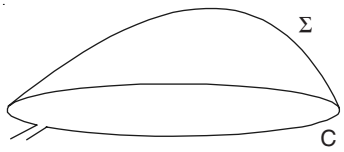


Figure 10.2 Induced emf in the wire loop.

$$\mathcal{E} = -\frac{d\Phi}{dt} \tag{10.13}$$

where Φ is the surface integral of **B** evaluated over any surface Σ with its edge on C.

The generalization

$$\mathcal{E} = -\frac{d\Lambda}{dt}$$

where Λ is the flux-linkage in a multi-turn coil which can also be shown to be related to a surface integral, though the surface is subtler. Thus, in words, the Faraday's law is:

> *The emf in the circuit is proportional to the rate of change of the flux linked by it, and the sense of the current flowing in the circuit is such that it (i.e. the current) opposes the change of the flux* Φ.

Now, the emf of a battery is the energy supplied per unit charge delivered; and an electric force **E** is set up within a wire connecting the terminals, and hence (Figure 10.3):

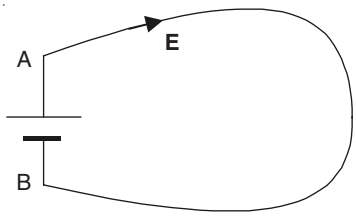


Figure 10.3 Battery loop.

The energy supplied per unit charge, $\mathcal{E} = \int_A^B E_l \, dl$

i.e. the line integral of **E** along the wire = $V_A - V_B$.

Hence the generalization of Eq. (10.13), for any loop whether of wire or otherwise is

$$\oint E_l \, dl = -\frac{d\Phi}{dt} \tag{10.14}$$

If we consider an elemental contour δC, enclosing an area δA, then

$$\Phi = \mathbf{B} \cdot \delta\mathbf{A}$$

∴ Equation (10.14) becomes

$$\oint_{\delta C} E_1\, dl = -\,\delta\mathbf{A}\cdot\frac{\partial\mathbf{B}}{\partial t} \qquad \text{or} \qquad \oint_{\delta C}\frac{E_1\, dl}{\delta\mathbf{A}} = -\,\frac{\partial\mathbf{B}}{\partial t} \tag{10.15}$$

As $\delta\mathbf{A} \to 0$, this equation becomes

$$\text{curl}\ \mathbf{E} = \nabla\times\mathbf{E} = -\,\frac{\partial\mathbf{B}}{\partial t} \tag{10.16}$$

i.e. at any point in a 'changing magnetic field', there is an associated electric field which is given by the above equation (which is one of the Maxwell's field equations).

Thus we have restricted ourselves to consider the stationary circuit subjected to a time-varying magnetic field only, at this stage.

Differential derivation

Unlike a closed surface (or a volume), a loop has a direction which is denoted by its normal, and hence essentially a vector. In Cartesian coordinate system, we can consider a general elemental loop by its projection on three coordinate planes. So first, we evaluate Eq. (10.14) for a rectangular loop ($\delta y \times \delta z$) lying in a plane perpendicular to the axis Ox (Figure 10.4).

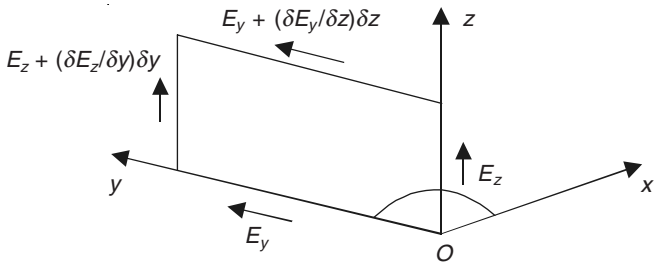


Figure 10.4 Rectangular loop in the *y-z* plane for evaluating the line-integral $E_1 dl$.

The variations of E_y and E_z for the rectangular loop are indicated in the above figure. So taking the line integral over the closed loop, we have

$$\oint \mathbf{E}\cdot d\mathbf{l} = E_y\,\delta y + \left(E_z + \frac{\partial E_z}{\partial y}\,\delta y\right)\delta z - \left(E_y + \frac{\partial E_y}{\partial z}\,\delta y\right)\delta y - Ez\,\delta z$$

$$= \left(\frac{\partial E_z}{\partial y}\right) - \left(\frac{\partial E_y}{\partial z}\right)\delta y\,\delta z$$

Also,

$$\Phi = B_x\delta y\,\delta z$$

Cancelling $\delta y\delta z$ from both the sides of Eq. (10.14), we get

$$\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = -\,\frac{\partial B_x}{\partial t} \tag{10.17}$$

By considering two more loops in the z–x and the x–y planes, two similar equations for the y- and z-directions will be obtained.

Note:

$\frac{d\Phi}{dt}$ = rate of change of Φ, with respect to time, in a particular circuit.

$\frac{\partial B_x}{\partial t}$ = time-rate of change of B_x, which also can vary with position.

We remind ourselves that the vector whose component in any direction **n** is the limit of the ratio

$$\frac{\text{Line-integral of } \mathbf{E} \text{ round a closed loop } C_\text{n} \text{ normal to } \mathbf{n}}{\text{Area of } C_\text{n}}, \text{ when } C_\text{n} \text{ is made very small}$$

is called the curl of **E** (= curl **E**).

Symbolically,

$$\left|\text{curl } \mathbf{E}\right| = \left|\nabla \times \mathbf{E}\right| = \lim_{\delta S_\text{n} \to 0} \frac{\oint_{C_\text{n}} \mathbf{E} \cdot d\mathbf{l}}{\delta S_\text{n}}$$

where δS_n is the area of C_n; and we have shown above (as well as earlier) that in Cartesian coordinates:

$$\text{curl } \mathbf{E} = \mathbf{u}_x \left(\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} \right) + \mathbf{u}_y \left(\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} \right) + \mathbf{u}_z \left(\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} \right)$$

(Note that the denominators are in cyclic order.)

The three equations of (10.17) are replaced by a single vector equation,

$$\text{curl } \mathbf{E} = \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$

which is same as Eq. (10.16). And, it should be noted that this is more general since it can be expressed in any coordinate system.

10.2.2 The EMF Induced in Conductors Moving in a Static Magnetic Field (The Flux Cutting Rule)

Faraday's law states the relationship between a time-varying magnetic field and its associated electric field. (It should be noted that a time-varying **E** or **B** cannot exist in isolation as a static **E** field—in a macroscopic sense). Also, the Faraday's law does not indicate what happens when a circuit moves through an *uniform magnetic field.* So, now let us consider a piece of metal wire (i.e. electrically conducting) moving with a velocity **v** in a static, arbitrary magnetic field **B**. All the elemental charges inside the wire would be subject to an electromagnetic force, given by

$$\mathbf{F} = Q\,(\mathbf{v} \times \mathbf{B})$$

Freely movable electrons in the wire would be pushed towards one end of the wire, making the other end positively charged (Figure 10.5).

In the static and the quasi-static case, the forces on the free charges cannot exist in a conductor. The distribution of the charges will be such that the **E** field created by these charges will be exactly equal and opposite to that of ($\mathbf{v} \times \mathbf{B}$). So, we can say that an external field **E** acts along the conductor, producing a total emf, i.e.

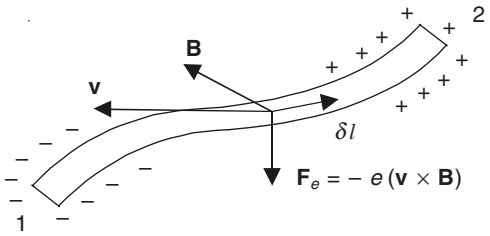


Figure 10.5 A piece of wire moving in a magnetic field.

$$\mathcal{E} = \int_1^2 (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{l} \qquad \text{(a moving piece of wire)} \tag{10.18}$$

Since the conducting path is not closed, there will be no current flowing in the conductor. But, when there is a closed conducting filament moving in a static magnetic field, the emf round the contour is given by

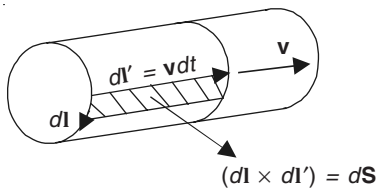


Figure 10.6 Geometry for the law of induction applied to a moving conductor.

$$\mathcal{E} = \oint_C (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{l} = \oint_C (d\mathbf{l} \times \mathbf{v}) \cdot \mathbf{B} \tag{10.19}$$

Let us say that dl' is the distance traversed by the contour in the time dt (Figure 10.6), so that $\mathbf{v} = d\mathbf{l}'/dt$. Now dt is constant, and since the integration is with respect to space, we can write Eq. (10.19) as

$$\mathcal{E} = \oint_C \left(d\mathbf{l} \times \frac{d\mathbf{l}'}{dt} \right) \cdot \mathbf{B} = \left(\frac{1}{dt} \right) \oint_C (d\mathbf{l} \times d\mathbf{l}') \cdot \mathbf{B}$$

$$= \left(\frac{1}{dt} \right) \iint_S \mathbf{B} \cdot d\mathbf{S}$$

(All the notations are as shown in Figure 10.6.)

If Φ is the magnetic flux at the initial instant, then the increase in the magnetic flux through the contour C in the time interval dt is

$$d\Phi = -\oint_C \mathbf{B} \cdot (d\mathbf{l} \times d\mathbf{l}') = -\iint_S \mathbf{B} \cdot d\mathbf{S}$$

$\therefore$ On substituting for the integral, we get

$$\mathcal{E} = -\frac{d\Phi}{dt} \qquad \text{(in a moving closed contour)} \tag{10.20}$$

The emf in this equation is often referred to as the 'motional emf'.

In any practical problem, when we deal with a closed circuit, we use the above equation, remembering that the magnetic flux is static and not changing with time, and the change of flux-linkage is due to the motion of the contour. When we deal with a moving piece of wire and not a closed contour, we use Eq. (10.18). For example, a body (a conductor of length l, say) which is moving in a steady magnetic field with a velocity $\mathbf{v}$, experiences an electric field $\mathbf{E} = (\mathbf{v} \times \mathbf{B})$; or a body in a magnetic field which is moving with a velocity $\mathbf{v}$ relative to the body, experiences a field given by $\mathbf{E} = -(\mathbf{v} \times \mathbf{B})$.

As an example, if a straight conductor of length l m is moving at a velocity of $\mathbf{v}$ m/s in a magnetic field $\mathbf{B}$, then

$$\mathcal{E} = \int_{\text{Cond}} \mathbf{E} \cdot d\mathbf{l} = \mathbf{l} \cdot (\mathbf{v} \times \mathbf{B})$$

If $\mathbf{l}$ is perpendicular to $\mathbf{B}$, then: $\mathcal{E} = Blv$.

10.3 THE GENERAL LAW OF ELECTROMAGNETIC INDUCTION

We now consider a 'completely general situation' in which a closed filamentary contour moved and deformed in an arbitrary manner in a time-varying magnetic field. This time-varying magnetic field would be associated with a time-varying electric field at all points and hence along the contour as well. Furthermore parts of the contour are moving in the magnetic field, and an electromagnetic force on the charges is also present. This part is then equivalent to an electric field intensity given by $(\mathbf{v} \times \mathbf{B})$. The total emf in the contour under this general condition is thus the sum total of Eqs. (10.12) and (10.19), i.e.

$$\begin{aligned} \mathcal{E} &= -\oint_C \left(\frac{\partial \mathbf{A}}{\partial t}\right) \cdot d\mathbf{l} + \oint_C (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{l} \\ &= -\iint_S \left(\frac{\partial \mathbf{B}}{\partial t}\right) \cdot d\mathbf{S} + \oint_C (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{l} \end{aligned} \tag{10.21}$$

which is the detailed form of Eq. (10.20).

The two terms on the right-hand-side of this equation can be expressed in a slightly similar form in terms of the flux Φ linked by the contour C. The first term is the (negative) rate of change of flux through the contour, due to the time-varying magnetic flux (when the contour is instantaneously held in position); and the second term is the (negative) rate of change of flux due to the motion and the deformation of the contour in the magnetic field assumed to be static at that instant. So we can rewrite the equation as

$$\mathcal{E} = -\left(\frac{\partial \Phi}{\partial t}\right)_{\text{varying field}} - \left(\frac{\partial \Phi}{\partial t}\right)_{\text{motion in the field}} \tag{10.22}$$

The first term is usually called the 'transformer emf', and the second term is known as the 'motional emf'. The above equation is same as the Eq. (10.20). This equation represents the 'Faraday's law of electromagnetic induction in general form'.

These equations should be applied to practical problems very carefully, because otherwise in a number of instances we can come to quite erroneous conclusions, as we shall see later.

10.4 AN ALTERNATIVE PROOF FOR THE GENERAL LAW OF ELECTROMAGNETIC INDUCTION (THE MOVING MEDIUM AND THE TIME-VARYING FIELD)

An alternative proof for the generalization of the Faraday's law is offered here in which we consider the most general situation whereby the magnetic field is time-varying and the circuit is moving with an arbitrary velocity **v** which need not be uniform, and the circuit itself is allowed to be deformable.

Furthermore this proof has the advantage that it does not require either the concept of the magnetic vector potential or the Lorentz force equation for the derivation. We merely need to use the Gauss' divergence theorem and the Stoke's theorem for our purpose. The proof was originally given by Corson and Lorrain.

We start from the integral form of the law of induction which was experimentally verified by Faraday for the stationary circuits, i.e.

$$\mathcal{E} = \oint_C \mathbf{E} \cdot d\mathbf{l} = -\frac{\partial}{\partial t} \iint_S \mathbf{B} \cdot d\mathbf{S}$$

which is our Eq. (10.12).

In the generalization, the source current creating the flux may be varying sinusoidally with time, so that **B** also varies the same way. So **B** can be expressed as

$$\mathbf{B} = \mathbf{B}\,(x, y, z, t)$$

i.e. **B** depends on both the space and the time coordinates. We consider a closed contour C which corresponds to the left-hand side of the above Eq. (10.12), and let each element $d\mathbf{l}$ of C move to a new position of C' (Figure 10.7) with a velocity **v** where **v** need not be uniform. Then the rate of change of the flux Φ is

$$\frac{d\Phi}{dt} = \frac{d}{dt} \iint_S \mathbf{B} \cdot d\mathbf{S}$$

$$= \lim_{\Delta t \to 0} \left[\frac{\iint_S \mathbf{B}'(t + \Delta t) \cdot d\mathbf{S}' - \iint_S \mathbf{B}(t) \cdot d\mathbf{S}}{\Delta t} \right] \tag{10.23}$$

where Δt is the travel-time, and $\mathbf{B}'(t + \Delta t)$ denotes the flux density across the surface S' at a time $(t + \Delta t)$. The volume swept out by the motion of C to C' is bounded by the surfaces S, S', and by the curved surface joining them (Figure 10.7).

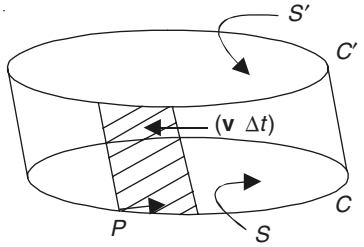


Figure 10.7 The path of integration moves from C to C' in the time Δt. The displacement is general (involving translation, rotation, and distortion).

Applying the Gauss' theorem, we get

$$\iiint \text{div } \mathbf{B}\, dv = \iint_{S'} \mathbf{B}'(t + \Delta t) \cdot d\mathbf{S}' + \int_C \mathbf{B}(t + \Delta t) \cdot (d\mathbf{l} \times \mathbf{v}\Delta t) - \iint_S \mathbf{B}(t + \Delta t) \cdot d\mathbf{S} \qquad (10.24)$$

where $(d\mathbf{l} \times \mathbf{v}\Delta t)$ is the shaded surface element shown in Figure 10.7, and the negative sign on the last integral indicates that the flux on the lower surface is opposite in sign to that in the upper one. Also it is to be noted that the instant under consideration is $(t + \Delta t)$.

By Taylor's theorem, the last integrand in Eq. (10.24) is

$$\mathbf{B}(t + \Delta t) = \mathbf{B}(t) + \frac{\partial \mathbf{B}}{\partial t} \Delta t + \qquad (10.25)$$

Applying the Stoke's theorem to the second integral on the right-hand side of Eq. (10.24), we get

$$\int_C \mathbf{B}(t + \Delta t) \cdot (d\mathbf{l} \times \mathbf{v}\Delta t) = \Delta t \int_C [\mathbf{v} \times \mathbf{B}(t + \Delta t)] \cdot d\mathbf{l}$$

$$= \Delta t \iint_S \text{curl}[\mathbf{v} \times \mathbf{B}(t + \Delta t)] \cdot d\mathbf{S} \qquad (10.26)$$

Substituting from Eqs. (10.25) and (10.26) in Eq. (10.24), and using the fact that $\nabla \cdot \mathbf{B} = 0$, we get

$$0 = \iint_{S'} \mathbf{B}'(t + \Delta t) \cdot d\mathbf{S}' + \Delta t \iint_S \text{curl}[\mathbf{v} \times \mathbf{B}(t + \Delta t) \cdot d\mathbf{S} - \iint_S \mathbf{B}(t) \cdot d\mathbf{S} + \frac{\partial}{\partial t} \iint_S [\mathbf{B}(t) \cdot d\mathbf{S}] \Delta t \qquad (10.27)$$

The superscript on S becomes unnecessary, as we calculate $(\Delta\Phi/\Delta t)$ in the limit. Hence from Eq. (10.23), we get

$$\frac{d\Phi}{dt} = \frac{\partial}{\partial t} \iint_{S'} \mathbf{B} \cdot d\mathbf{S} - \iint_S \text{curl } (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{S} \qquad (10.28)$$

and the Faraday's law of induction generalizes to

$$\mathcal{E} = \oint_C \mathbf{E} \cdot d\mathbf{I} = -\frac{\partial}{\partial t} \iint_{S'} \mathbf{B} \cdot d\mathbf{S} - \iint_S \text{curl } (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{S} \qquad (10.29)$$

which is same as Eq. (10.21).

10.5 EXAMPLES OF APPLICATION OF ELECTROMAGNETIC INDUCTION

10.5.1 Transformer

A transformer essentially consists of two windings on a closed iron circuit (called the core, which is made up of 3% Si–Fe laminations). It is schematically represented, in its simplest form as shown in the Figure 10.8 (not pictorially).

For the present analysis, we assume both the windings to have zero resistance. When an emf $\mathcal{E}_1$ is applied to the primary winding of N_1 turns, a current passes through it and sets up a flux Φ in the iron core. The emf induced by this flux must be exactly equal and opposite to $\mathcal{E}_1$,

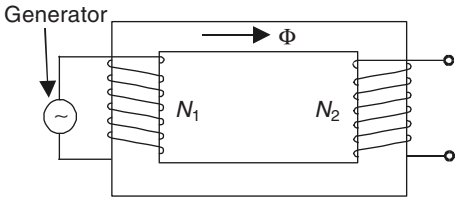


Figure 10.8 A schematic representation of a single-phase two-limb two-winding transformer.

because if this did not happen, then there would be a residual voltage in the winding which would set up an infinite current in the zero-resistance primary circuit. Mathematically $\mathcal{E}_1$ and Φ are related by the equation

$$\mathcal{E}_1 = N_1 \frac{d\Phi}{dt} \tag{10.30}$$

and the current drawn by this primary winding is of that exact magnitude which is required to set up this flux Φ in the core. The same flux also links the secondary winding on the other limb of the core. If this winding has N_2 turns, then the emf $\mathcal{E}_2$ set up in this winding is

$$\mathcal{E}_2 = N_2\left(\frac{d\Phi}{dt}\right) = \left(\frac{N_2}{N_1}\right)\mathcal{E}_1 \tag{10.31}$$

Thus the variations of $\mathcal{E}_1$ are reproduced in $\mathcal{E}_2$, and are scaled by the turns ratio (N_2/N_1).

Equation (10.31) cannot hold under all conditions; for example, let $\mathcal{E}_1$ be a constant direct voltage (i.e. of steady value) suddenly switched on and allowed to persist for a long duration of time. In such a case, a constant $\mathcal{E}_1$ would imply a constant $(d\Phi/dt)$, and since $\mathcal{E}_1$ persists for a long time, then Φ would have to increase indefinitely without limit, which in reality cannot be realized because of magnetic saturation of iron. As the saturation flux is approached, the magnetizing current would increase rapidly, and this would cause, in the winding, a voltage drop across its resistance too great to be neglected, thus violating our initial assumption. Hence the transformers will not transform direct voltages, and can *only* be used for transforming alternating voltages. A primary voltage which varies sinusoidally with time, would be very faithfully reproduced across the terminals of the secondary; but other shaped waves, such as triangular, rectangular, saw-tooth, etc. can also be transmitted by transformers of suitable design.

Suppose, now we connect a load (electrical) across the secondary terminals which draws a current I_2 (Figure 10.9). This current causes an mmf (N_2I_2) to the magnetic circuit of the core,

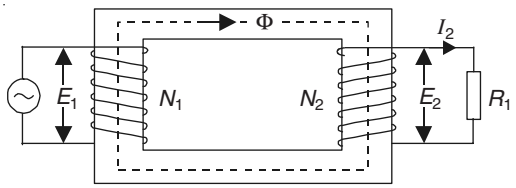


Figure 10.9 A transformer drawing a load current.

tending to change the magnetic flux wave. But the changes in the flux wave are controlled by any changes in the emf, governed by Eq. (10.30). So the effect of I_2 has to be counterbalanced by an extra current I_1' in the primary winding (drawn from the supply source), such that

$$N_1I_1' + N_2I_2 = 0 \tag{10.32}$$

The total current in the primary winding is I_1 which is the sum of I_1' and the magnetizing current (= I_m) required to set up the flux in the core, when the secondary winding draws no current (i.e. is in open-circuit). In power transformers $I_1' >> I_m$, and hence:

$$I_1 = I_1' \text{ approximately}$$
$$= -\left(\frac{N_2}{N_1}\right) I_2 \tag{10.33}$$

That is, the ratio of the primary to the secondary current is (nearly) equal to the turns ratio and the two currents tend to magnetize the core in opposite senses.

The transformer is a device in which the law of electromagnetic induction has to be applied in the form of 'flux-linking'.

10.5.2 Direct Current Generator (Heteropolar Machine)

Most types of electric generators are so designed that they produce their induced emf by a relative movement between a set of coils, which are mounted in iron, and a set of magnetic field producing poles. In the dc generator, the armature coils are carried in the slots on the rotor (armature), as shown in the Figure 10.10(a).

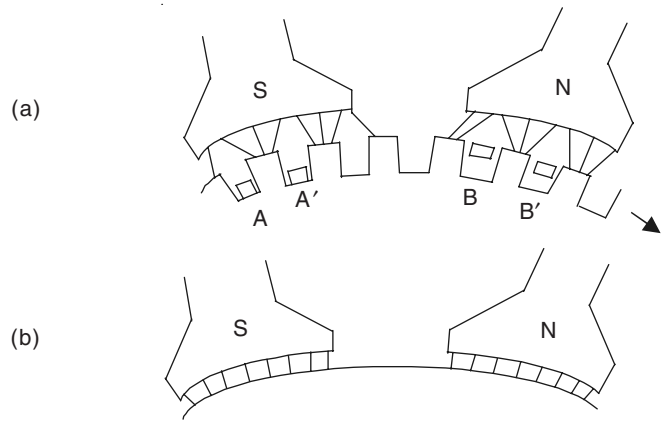


Figure 10.10 Section of a dc generator.

Figure 10.10(a) shows a coil with the sides in the slots marked A and B, the width of the coil being nearly the same as the distance between the centre-lines of the successive poles which are of opposite polarity (hence the name heteropolar machine). Since we are looking at a single coil AB at the instant shown in the Figure 10.10(a), the flux linked by the coil is the resultant of the flux entering from the *N* pole and the flux going out under the *S* pole. In a time-interval *t*, the rotor turns through one tooth pitch and the coil now occupies the position A′B′, and the flux linked by AB has changed because of the increase in the flux between BB′ and a further decrease in the flux between A′A. If *l* is the length of the coil-sides, then the increase in the flux linked is ($2Bl$AA′) or ($2Bl$BB′), and

$$\text{AA}' = \text{BB}' = \text{v}t$$

where v is the velocity of the rotor, and this gives the emf induced in the coil AB as

$$\mathcal{E}_1 = 2Bl\text{v} \tag{10.34}$$

where B is the mean flux density normal to the rotor surface. So we can say that each coil-side has contributed

$$\mathcal{E} = Blv \tag{10.35}$$

to the induced emf. It should be noted that we have used the method of flux-linkage, even though the emf has been produced by flux cutting. We could have used the method of flux cutting, if we assumed a smooth rotor surface and the coil-sides placed touching the rotor surface, as a simplified idealization of the problem [Figure 10.10(b)]. The flux density under each pole is then uniform, and we find that each coil-side, by the method of flux-cutting, has the induced emf as Blv. We have shown this for a single coil per pole pair. If, now, the machine has $2p$ poles, each with a flux Φ, and the speed of rotation is n revolutions per second, then the mean emf per conductor is

$$\bar{\mathcal{E}}_1 = 2pn\Phi \tag{10.36}$$

If the armature has Z conductors so connected that there are a parallel paths, then the mean emf for the whole machine is

$$\bar{\mathcal{E}} = \left(\frac{Z}{a}\right)\bar{\mathcal{E}}_1 = \frac{2pn\Phi Z}{a} \tag{10.37}$$

It should be noted that as the coils move across the successive magnetic poles, which are of opposite polarity, the induced emf would also be of alternating polarity, and hence a mechanical switch is used to reverse its polarity every half-cycle, so that in the external circuit a unidirectional current flows. Such a switch is called a *commutator*, and we shall not go into the details of such components, which can be found in any textbook on electrical machines.

10.5.3 Search Coil

This is a very important device for measuring the magnetic flux and the flux density (both due to the direct current as well as the alternating current). It uses a small coil of many turns, N, each of mean area A, placed so that the required **B** crosses it normally. The coil is made up of very thin wire (usually 30 to 40 SWG enameled copper wire), and is positioned in the magnetic field to be measured, assuming that the coil dimensions are much smaller than the curvature of the field lines. The ends of the coil are usually tightly twisted together, and connected to the measuring device, which can be a ballistic galvanometer or an integrating fluxmeter (for direct current fields). When the field B lines cross the coil normally, the flux-linkage is then NAB (or $NAB \cos \theta$, if the direction of **B** makes an angle θ with the normal to the plane of the search coil). A flux change of NAB (or $NAB \cos \theta$) or $2NAB$ (or $2NAB \cos \theta$) is produced by removing the coil to a place where $\mathbf{B} = 0$, or by reversing the direction of B, whichever is quicker and more convenient. As mentioned above, this flux change is measured by an integrating fluxmeter, or the change of charge Q

$$= \int_{t_i}^{t_f} I\,dt = -\frac{1}{R}\int_{\Phi_i}^{\Phi_f} d\Phi = \frac{\delta\Phi}{R}$$

by a ballistic galvanometer, provided the change has been made quickly enough. This is the technique for measuring the direct current fields. For measuring the alternating flux, the search coil when placed in the field, produces an induced emf which is proportional to the frequency and the flux linked by the coil, i.e.

$$\mathcal{E} = NAB\omega \sin \omega t = \Phi_{\max}\, \omega \sin \omega t \tag{10.38}$$

For a detailed analysis of the behaviour of search coils, readers are recommended papers by S.C. Bhargava, and A.D. Moore. By arranging and winding 2 or 3 search coils together in orthogonal planes on suitably designed formers, it is possible to build a composite coil which would measure all the components of 2- or 3-dimensional fields, simultaneously, at any given point. Figure 10.11 shows a typical search coil.

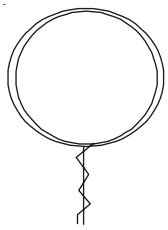


Figure 10.11 A search coil.

10.6 INDUCTANCE

Self-inductance of a coil is defined as the flux-linkage of the coil per ampere of the current flowing in it, i.e.

$$L = \frac{\Phi}{I} \tag{10.39}$$

where Φ is the flux linked by the coil when it is carrying a current I.

Mutual inductance. When a current flows in a circuit, a magnetic flux is set up. If a part of this flux is linked with a second circuit, then there is said to be a mutual inductance between the two circuits. The mutual inductance between the circuits (1) and (2) as shown in Figure 10.12 is defined as

$$M_{21} = \frac{\Phi_2}{I_1} \tag{10.40}$$

i.e. the flux-linkage in the circuit (2) due to a unit current in the circuit (1).

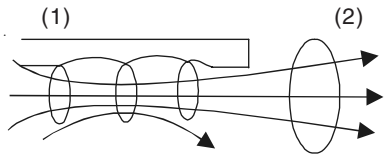


Figure 10.12 Mutual inductance between two circuits.

If the roles of the two circuits are reversed, then

$$M_{12} = \frac{\Phi_1}{I_2} \tag{10.41}$$

Let us consider a simple, two-limb, single-phase transformer (Figure 10.13). In the transformer, the primary current I_1 magnetizes the core whose permeance, let us say, is Λ.

$\therefore$ The flux set up in the core is

$$\Phi = \Lambda N_1 I_1$$

where the primary winding has N_1 turns.

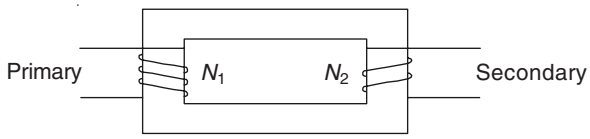


Figure 10.13 Single-phase transformer.

If the secondary winding has N_2 turns, then the mutual inductance between the primary and the secondary is

$$M_{21} = \Lambda N_1 N_2 \tag{10.42}$$

Symmetry of the situation shows that if the secondary was excited, the mutual inductance for the primary would be

$$M_{12} = \Lambda N_2 N_1 \tag{10.43}$$

i.e.

$$M_{12} = M_{21}$$

The proviso in this equality is that the permeances of both the paths are the same (i.e. constant permeance). This is a restraint on iron-cored circuits, but holds more generally for all air-cored circuits, i.e. $M_{12} = M_{21}$, though this is not so easily obvious in all cases. An example is the positioning of a short coil inside a long coil (solenoid), as shown in Figure 10.14. As is seen in

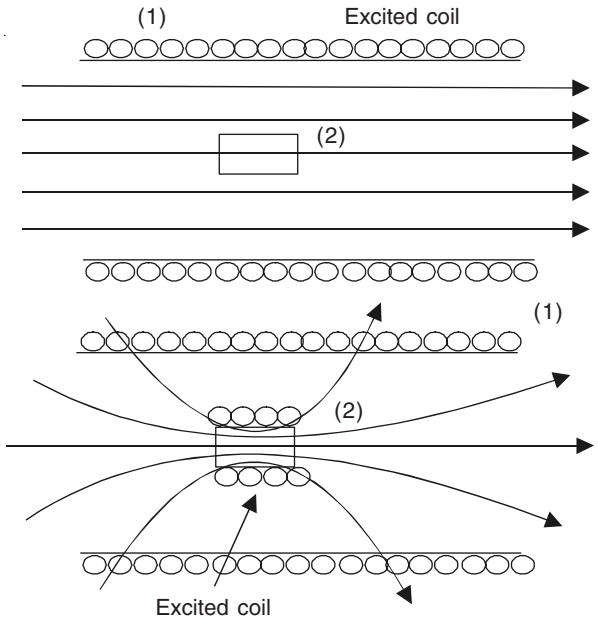


Figure 10.14 A short coil inside a long coil, for the equivalence of M_{12} and M_{21}.

this figure, calculating the flux-linkage in the circuit (2) due to a current in (1) is a simple matter, but the vice versa is quite complex.

Unit of inductance. The inductance is 1 henry if the passage of 1 ampere of current in circuit (1) sets up a flux-linkage of 1 weber-turn in the circuit (2).

Note: Permeance = Reciprocal of Reluctance (= R).

In a magnetic circuit, the magnetic resistance or the reluctance is

$$\text{Reluctance} = \frac{\text{MMF}}{\text{Flux}}$$

Its unit is amp-turn/weber.

10.6.1 Inductance in Terms of Induced Emfs

When the current is changing in one circuit, it sets up a changing flux and hence an induced emf in the other circuit. The same argument applies to a single circuit.

First, we consider a single air-cored circuit, for which we have

$$\Phi = LI$$

$$\therefore \quad \text{The induced emf} = \mathcal{E} = \frac{d\Phi}{dt} = L\frac{dI}{dt} \tag{10.44}$$

And for an iron-cored circuit,

$$\text{the induced emf} = \mathcal{E} = \frac{d\Phi}{dt} = L\frac{dI}{dt} + I\frac{dL}{dt}$$

$$= \left(L + I\frac{dL}{dI}\right)\frac{dI}{dt} \tag{10.45a}$$

since the inductance L of an iron-cored circuit depends on the current as well.

Next, we consider circuits with mutual inductance, i.e.

$$\Phi_2 = M_{21}I_1 = M_{12}I_1$$

$$\therefore \quad \mathcal{E}_2 = M_{12}\left(\frac{dI_1}{dt}\right)$$

and, similarly for the other circuit,

$$\mathcal{E}_1 = M_{12}\left(\frac{dI_2}{dt}\right) \tag{10.45b}$$

The mutual inductance is 1 henry, when $\mathcal{E}$ = 1 volt, and the current in the other coil is changing at 1 amp/s.

When the circuit is iron-cored, M_{12} changes with the current as well, and hence

$$\mathcal{E}_2 = \frac{d}{dt}(M_{12}I_1) = M_{12}\frac{dI_1}{dt} + I_1\left(\frac{dM_{12}}{dt}\right)$$

If $I_2 = 0$, or M_{12} varies due to the changes in I_1, then only

$$\mathcal{E}_2 = \left(M_{12} + I_1\frac{dM_{12}}{dI_1}\right)\left(\frac{dI_1}{dt}\right)$$

and, for a similar behaviour,

$$\mathcal{E}_1 = \left(M_{12} + I_2\frac{dM_{12}}{dI_2}\right)\left(\frac{dI_2}{dt}\right) \tag{10.46}$$

Mutual inductors in which the coils are of toroidal type, as discussed in Section 7.10.5, are of great use as measuring circuits because they can be made with great accuracy. The great advantage of such an inductor is that it is practically unaffected by the presence of any neighbouring circuits. A coil of this type, used for measuring alternating currents non-invasively, is known as 'Rogowski coil'.

10.6.2 Calculation of Inductances

The definitions of self- and mutual-inductances are subject to certain uncertainties, especially the self-inductances. This is because of the difficulties in defining exactly the flux linkage. To illustrate this uncertainty caused, we consider the following two circuits. Figure 10.15(a) shows

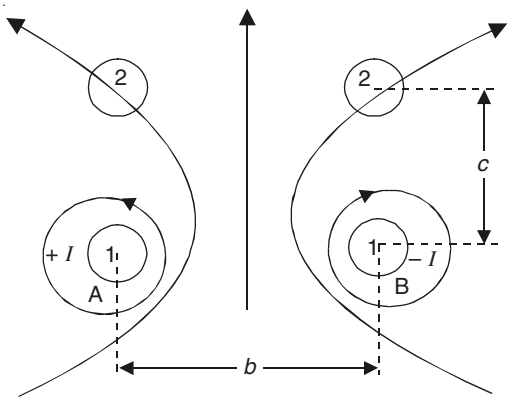


Figure 10.15(a) Mutual inductance between parallel conductors of finite cross-section.

the cross-section of the two circuits, each consisting of a pair of thick parallel wires. Some of the flux produced by the currents in the circuit (1), traverse the cross-section of the wires of the circuit (2). The question arises as to whether this flux is linking with the circuit or not. In this example, which is that of calculating the mutual inductance, the difficulty is evaded by assuming the wires to be infinitely thin. The total flux linked by the circuit (2) can be calculated by superimposing the effects of the two conductors of the circuit (1) on the circuit (2). Hence using the dimensions as shown in the figure the flux linked by (2) due to the current I in A alone is

$$\frac{\mu_0 I}{2\pi}\int_{c}^{\sqrt{b^2+c^2}}\frac{dr}{r}=\left(\frac{\mu_0 I}{4\pi}\right)\ln\left(1+\frac{b^2}{c^2}\right) \quad \text{webers/metre}$$

noting that, due to an infinite conductor, $\mathbf{B} = \dfrac{\mu_0 I}{2\pi r}$.

The conductor B also contributes an equal quantity, and the flux due to both the conductors of the circuit are in the same direction.

$$\therefore \qquad M_{12}=\frac{\Phi_2\ (\text{total})}{I_1}=\left(\frac{\mu_0}{2\pi}\right)\ln\left(1+\frac{b^2}{c^2}\right) \quad \text{henries/metre} \qquad (10.47)$$

This neglects the finite cross-section of the conductors, but the result is accurate enough for most of the purposes.

We consider the second example, as shown in Figure 10.15(b) in which we look at the self-inductance of the circuit. Here because of the finite cross-section of the conductors, some of the

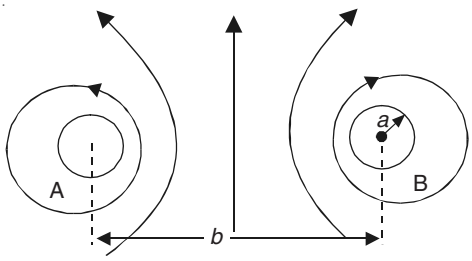


Figure 10.15(b) Self-inductance of parallel conductor circuit.

flux is actually within the wires. The question now is, should this flux be taken as linking the circuit? If, as in the previous problem, we neglect the conductor dimensions, then:

$$\text{The flux } \Phi \text{ due to the current } I \text{ in A} = \frac{\mu_0 I}{2\pi}\int_0^b \frac{dr}{r}$$

$$= \left(\frac{\mu_0 I}{2\pi}\right)\ln r\Big|_0^b$$

and $\ln 0 \to \infty$.

Hence the wire diameter cannot be neglected. Also we should note that finer the wire, the higher the inductance. So if the wire has a radius a, then

$$\Phi \text{ due to the current } I \text{ in A} = \frac{\mu_0 I}{2\pi}\int_a^{b-a} \frac{dr}{r}$$

$$= \left(\frac{\mu_0 I}{2\pi}\right)\ln\left(\frac{b-a}{a}\right)$$

$$\therefore \quad L \text{ due to the currents in A and B} = \left(\frac{\mu_0}{\pi}\right)\ln\left(\frac{b-a}{a}\right)$$

$$= \left(\frac{\mu_0}{\pi}\right)\ln\left(\frac{b}{a}\right) \tag{10.48}$$

This formula is reasonably accurate for the low frequency hollow conductors and high frequency solid conductors (due to the skin effect to be studied later in the Chapter 15). However this present calculation neglects the flux inside the solid conductor. Later we shall see how this contribution to the inductance (in Chapter 13) can be taken into account.

Now we derive the expression for the mutual inductance of the arrangement shown in the Figure 10.14. Let the current in the longer solenoid be I, and let it have N_1 turns per unit length, and the smaller solenoid have N_2 turns per unit length, its axial length being l. The flux density produced by the 'longer' solenoid = $\mu_0 N_1 I$.

$\therefore$ The flux linked by the smaller solenoid = $\mu_0 N_1 N_2 I l A$, where A is the cross-sectional area of the smaller solenoid.

$$\therefore \qquad M = \mu_0 N_1 N_2 l A \tag{10.49}$$

Next, we consider another problem, i.e. a small circular coil on the axis of a large circular

coil (Figure 10.16). If the number of turns of the large coil is N_1 and its radius is a_1 and carries a current I, then its magnetic flux density along its axis at a distance x from the plane of the coil is

$$B = \frac{\mu_0 N_1 I a_1^2}{2\left(a_1^2 + x^2\right)^{3/2}}$$

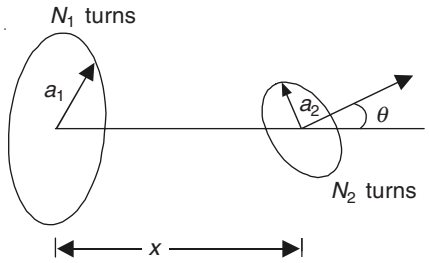


Figure 10.16 A small circular coil on the axis of a large circular coil.

Assuming the cross-section of the small coil to be small enough, the flux linked by it is

$$= N_2 \pi a_2^2 B$$

and hence

$$M = \frac{\mu_0 N_1 N_2 \pi a_1^2 a_2^2 \cos\theta}{2\left(a_1^2 + x^2\right)^{3/2}} \tag{10.50}$$

where θ is the angle between the axes of the two coils.

Next, we consider the self-inductance of a toroid (Figure 10.17) which is made up of an

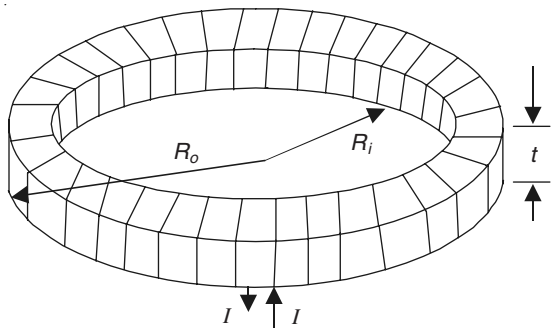


Figure 10.17 Inductance of a toroidal coil.

N-turn coil tightly wound around a dough-nut shaped core of rectangular cross-section with the inner and the outer radii of R_i and R_o, respectively, and the core having the permeability μ. We solve this problem (for the initial steps) by applying the Ampere's law to find the flux density. The only contour we need to consider is the one with its radius lying between R_i and R_o, since all other contours, i.e. $r < R_i$ and $r > R_o$ would link no current, and hence be equal to zero.

$\therefore$ For a contour such that $R_i < r < R_o$,

$$\oint \mathbf{H} \cdot d\mathbf{l} = H_\phi 2\pi r = NI$$

The flux through a single loop,

$$\Phi = \mu t \int_{R_i}^{R_o} \left(\frac{NI}{2\pi r} \right) dr$$

$$= \left(\frac{\mu t N I}{2\pi} \right) \ln \left(\frac{R_o}{R_i} \right)$$

$\therefore$ Self-inductance, $L = \left(\frac{\mu t N^2}{2\pi} \right) \ln \left(\frac{R_o}{R_i} \right)$ (10.51)

(the term N^2 appears in this expression because there are N turns in all of the coil; and t is the axial thickness of the toroid).

It is this type of toroid which is generally used for Rogowski coils, though to evaluate its induced emf, we need to consider its mutual inductance when a straight conductor passes through the centre of the toroid core in the direction normal to its plane.

10.6.3 Relation between the Self- and the Mutual-inductances of Two Coils (The Coefficient of Coupling)

We now investigate the relationship between the self-inductances of two coils (say L_{11} and L_{22}), and their mutual-inductance M_{12} (Figure 10.18). Let the number of turns of the two coils be N_1 and N_2, respectively.

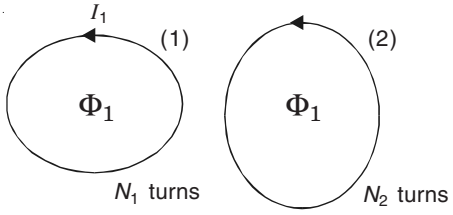


Figure 10.18 Mutual inductance between two coils.

When the two coils are very closely coupled, i.e. there is no leakage flux, Φ_1 being the flux linked per turn due to the current I_1 in the circuit (1), then

$$L_{11} = \frac{N_1 \Phi_1}{I_1} \quad \text{and} \quad M_{12} = \frac{N_2 \Phi_1}{I_1}$$

$\therefore$ $$\frac{L_{11}}{M_{12}} = \frac{N_1}{N_2}$$

Similarly for a current (say I_2) in the circuit (2), we get the relationship

$$\frac{L_{22}}{M_{21}} = \frac{N_2}{N_1}$$

and since $M_{12} = M_{21}$

$\therefore$ $$M_{12}^2 = L_{11} L_{22} \quad \text{or} \quad M_{12} = \pm \sqrt{L_{11} L_{22}}$$ (10.52)

If the positive directions of the current-flow in the two coils are so chosen as to produce the fluxes in the same direction, then the value of M_{12} will be positive, otherwise negative.

If the two coils are loosely coupled, i.e. there is leakage flux, then

$$L_{11} = \frac{N_1 \Phi_1}{I_1} \quad \text{and} \quad M_{12} = \frac{N_2 k_1 \Phi_1}{I_1}, \; k_1 < 1$$

$$\therefore \quad \frac{L_{11}}{M_{12}} = \frac{N_1}{k_1 N_2}$$

Similarly for the second coil,

$$\frac{L_{22}}{M_{12}} = \frac{N_2}{k_2 N_1}, \; k_2 < 1$$

$$\therefore \quad M_{12}^2 = k_1 k_2 L_{11} L_{22} \quad \text{or} \quad M_{12} = \pm k\sqrt{L_{11} L_{22}}$$

$$k^2 = k_1 k_2, \;\; 0 < k < 1 \tag{10.52a}$$

where k is called the coupling factor of the coils.

If the two coils are connected in series with the fluxes supporting each other, then

$$\mathcal{E}_1 = L_{11}\left(\frac{dI}{dt}\right) + M_{12}\left(\frac{dI}{dt}\right) \quad \text{and} \quad \mathcal{E}_2 = L_{22}\left(\frac{dI}{dt}\right) + M_{12}\left(\frac{dI}{dt}\right)$$

$$\therefore \quad \mathcal{E}_1 + \mathcal{E}_2 = (L_{11} + L_{22} + 2M_{12})\left(\frac{dI}{dt}\right) \tag{10.53}$$

$\therefore$ The combined self-inductance of the circuit,

$$L = L_{11} + L_{22} + 2M_{12}$$

If the coils are connected so that the fluxes oppose each other, then

$$L = L_{11} + L_{22} - 2M_{12} \tag{10.54}$$

10.7 SOME DIFFICULTIES IN INTERPRETING THE LAWS OF ELECTROMAGNETIC INDUCTION

So far we have used the laws of induction to solve a number of problems. But there are a number of cases where there are difficulties in interpreting these laws correctly. These difficulties arise more particularly in those problems which combine both the components of the induced emf, i.e. the transformer emf and the motional emf. We shall discuss some of such problems at this stage, though some more problems will be left for a later stage when we consider the laws of electromagnetism as seen by an observer in relative motion (i.e. moving media problems based on 'electromagnetism and special relativity'). However the first problem, which we shall now consider arises out of the difficulties in defining the concept of the flux-linkage in an unambiguous manner.

10.7.1 Problems due to the Definition of Flux-linkage

The definition of flux-linkage as 'the sum of the fluxes through all the turns of a circuit' is

neither perfect nor workable in all cases. Let us consider an air-cored solenoid with widely spaced turns (Figure 10.19). In this device, a 'turn' is not a closed loop, and hence counting the

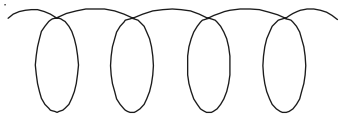


Figure 10.19 An air-cored solenoid with widely-spaced turns.

number of tubes of flux linking a particular turn is somewhat ambiguous. This uncertainty can be partly eliminated, because if a tube of flux is a closed ring then there is no ambiguity, but if the tube of flux is helical in nature, then the difficulty crops up again. Basically the operation of calculating is as follows. We form an open surface S with its edge on the contour C (Figure 10.20). The surface S is divided into small elements δS, and at each element, we evaluate

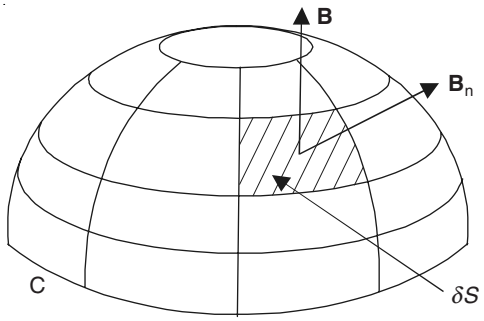


Figure 10.20 Open surface S.

$\mathbf{B} \cdot d\mathbf{S}$ (i.e. the quantity of the normal component of $\mathbf{B}$ through the surface element $d\mathbf{S}$). The total flux of $\mathbf{B}$ is then given by

$$\Phi = \Sigma\, \mathbf{B} \cdot d\mathbf{S}\ (= \Sigma\, B_n dS) = \iint_S \mathbf{B} \cdot d\mathbf{S} \tag{10.55}$$

Mathematically, this definition is quite unambiguous, but the difficulty arises due to the choice of the surface S, e.g. for a two-turn helix as shown in Figure 10.21. The circuit in this case is

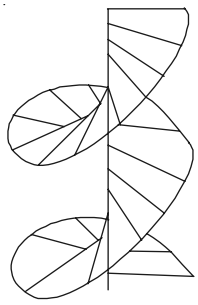


Figure 10.21 A two-turn helix.

closed by joining its two ends by parallel radial lines to a central axial line. On such a surface as shown in this figure, some tubes of flux can contribute twice to the integral, whilst some may contribute only once. Care must also be taken of the proper sign.

We shall next show another multi-turn contour which gives rise to two possible surfaces as shown in Figure 10.22. The question here is: in the figure shown, which of the two alternative

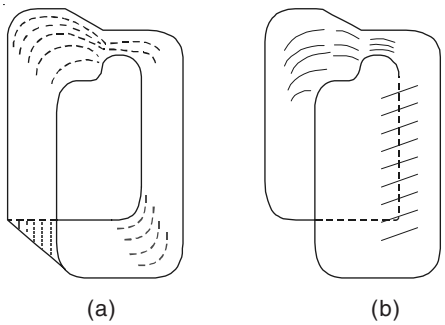


Figure 10.22 Two alternative surfaces spanning a two-turn loop.

surfaces can be used for finding the induced emf due to a time-varying magnetic field? Here we have to find the surface on which we would find the flux-linkage. The correct answer is the surface (b), since it is two-sided or orientable. The surface (a) is one-sided, Möbius surface, and cannot be used for flux-linkage. This, in fact, is the general problem for all multi-turn loops.

The distinction between the exact and the approximate definitions of the flux-linkage is further shown by the following example. An insulating cylinder C located in a uniform axial magnetic field **B** (as shown in Figure 10.23) rotates uniformly so as to wind wire from a

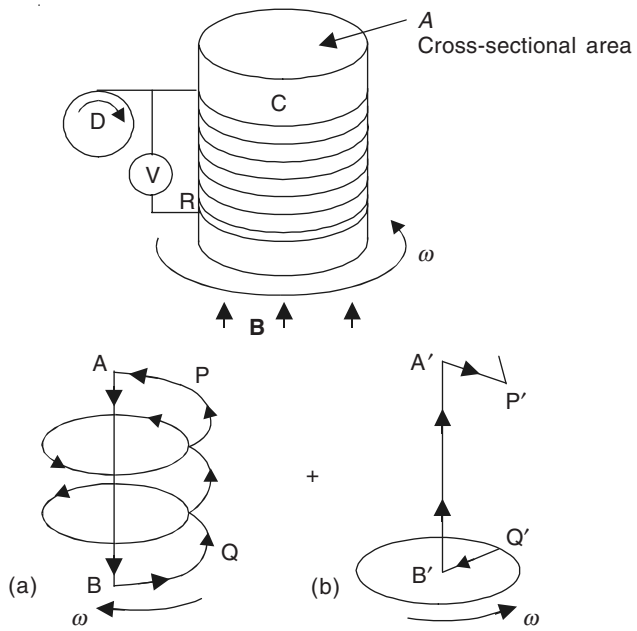


Figure 10.23 Time-varying number of turns on a coil in a uniform steady magnetic field.

drum D. The end of the wire is anchored to a contact ring R fixed to the lower end of the cylinder. The question here is: would there be an induced emf between R and the end of the incoming wire, as there is an apparent increase in the flux-linkage due to the increasing number of turns?

In fact, no emf is induced, and the voltmeter V does not show any reading. The exact definition of the flux-linkage explains this result. In this case, the emf induced is the resultant of the emfs in the two contributory circuits as shown above:

(a) a lengthening spiral (helix) joined to its central axis by a radius AP which is fixed in direction and a radius BQ which rotates; and
(b) a shaft A′B′, joined to the external connections by a radius A′P′ which is fixed in direction, and a radius B′Q′ which rotates as a spoke of the contact ring.

Each time a turn is added, the linkage in the circuit (a) is increased by the flux through the extra convolution, but equal flux is swept through by the moving radius of the circuit (b), and the connections of the two are such that the induced emfs are in opposition, leaving zero resultant. The superficial argument takes account of (a) only, and hence leads to a wrong result.

So for a constant flux in the cylinder, i.e. Φ = (BA) per turn, even though the number of turns are changing, no flux is generated, i.e.

$$\mathcal{E} = N(t)\left(\frac{d\Phi}{dt}\right)$$

If Φ is constant, independent of time, then even with increasing $N(t)$ (i.e. a function of time), $\mathcal{E} = 0$.

If, however, Φ itself is proportional to the number of turns, i.e. a direct current is being fed into the increasing number of turns, then: $\Phi = \mu_0 N(t) IA/l$, where l is the axial length of the wound cylinder, and A is its cross-sectional area, i.e. an emf would be generated (Figure 10.24).

$$\mathcal{E} = N(t)\left(\frac{d\Phi}{dt}\right) = \left(\frac{\mu_0 N(t) A}{l}\right)\frac{d}{dt}\left[N(t) I\right] \tag{10.56}$$

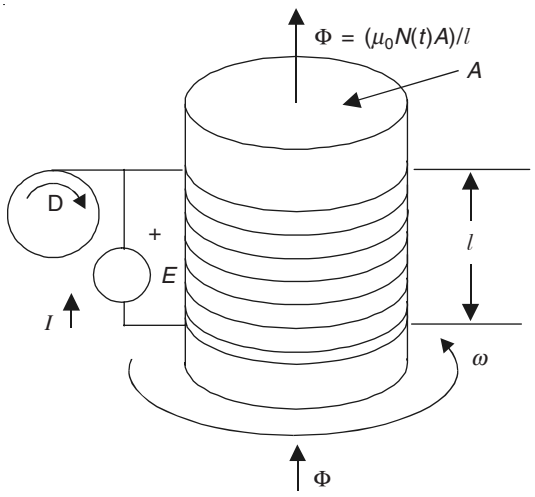


Figure 10.24 Time-varying number of turns on a coil in a uniform increasing magnetic field.

If I is a direct current, then the differentiable term is $I\frac{d}{dt}N(t)$, and hence even for direct current, there will be an induced emf. Note also that there is no motional emf contribution in this problem.

10.7.2 Problem Arising out of the Concept of 'Moving Field'

The concept of the moving magnetic field is in great use by the designers of electrical machines, e.g. the magnetic field set up by the rotating pole system in the synchronous generators or by a set of distributed three-phase current system in induction machines. In these devices, the 'moving field' is interpreted as a 'moving flux density pattern'—a flux configuration that glides along, but does not change. This is a legitimate concept, and as such the idea of the moving field can be used without any ambiguity. But there are times when such ideas can be misinterpreted. For example, let us consider a cylindrical bar magnet (which is essentially an axi-symmetric object) made to rotate about its axis, and a circuit is made by making contacts with its axis (i.e. the shaft of the magnet) and with its equator, as shown in Figure 10.25. When these sliding contacts are

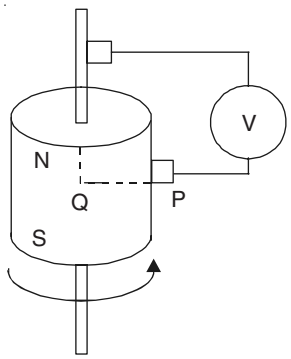


Figure 10.25 A rotating cylindrical bar magnet.

made, an induced emf is observed in the circuit. The question that has been raised in this experiment is: "do the lines of force rotate with the magnet inducing an emf in the circuit as they (the lines of force) cut the external circuit" or "do they stand still, inducing the emf as they are cut by the material of the magnet"? The question is, in fact, a meaningless one (as explained by Prof. G.W.O. Howe) because since the pattern is an axi-symmetric one, there is absolutely no change in the magnetic condition of the surrounding space when the magnet rotates. It is meaningless to inquire whether the lines by which we represent this condition (and these lines in fact are a figment of our imagination) have rotated or not. So the right answer can be obtained by either assumption. Some experts (among whom there was Prof. G.W. Carter) prefer the second option because this is then a case of direct application of the 'Neumann's law'.

10.7.3 The Puzzle of the Commutatorless DC Generator

The following device, which was suggested as a dc generator without moving contacts (which has been a dream of many cranks in electrical engineering), is another example of the confused application of the law of induction. The basic principle of operation of the device is as follows, explained in Figure 10.26. A wire of length l (along the x-direction) oscillates in the z-direction with a velocity $\mathbf{v}$ given by

$$\mathbf{v} = \mathbf{i}_z v_m \cos \omega t \tag{10.57}$$

in a magnetic field $\mathbf{B}$ which is also a function of time with the same frequency, i.e.

$$\mathbf{B} = \mathbf{i}_y B_m \cos \omega t \tag{10.58}$$

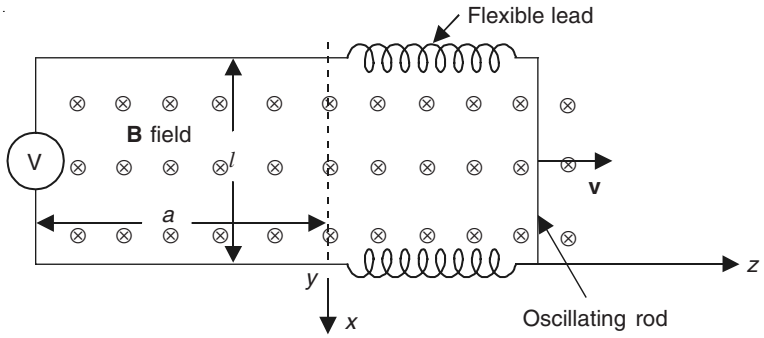


Figure 10.26 Principle of operation of the commutatorless dc generator.

The inventor suggested this to be a dc generator without moving contacts, because the induced emf is (by the flux cutting rule):

$$\mathcal{E} = Blv = B_m v_m l \cos^2 \omega t = \left(\frac{1}{2}\right) B_m v_m l(1 + \cos 2\omega t) \tag{10.59}$$

i.e. the generator produces a steady voltage, plus a double-frequency component which can be filtered out.

This argument is wrong, because the above component is really based on the part ($\mathbf{v} \times \mathbf{B}$), i.e. the motional emf component only and completely ignores the transformer component.

So we use the general induction theorem:

$$\oint_C \mathbf{E} \cdot d\mathbf{l} = -\iint_S \left(\frac{\partial \mathbf{B}}{\partial t}\right) \cdot d\mathbf{S} + \oint_C (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{l} \tag{10.60}$$

where the second term has been converted into a line integral by the Stoke's theorem, for the convenience of calculation. We consider the contribution to the emf due to the motion of the rod, i.e.

$$\oint (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{l} = \int_0^l (\mathbf{i}_z v_m \cos \omega t \times \mathbf{i}_y B_m \cos \omega t) \cdot \mathbf{i}_y dl$$

$$= -v_m B_m l \cos^2 \omega t = -\frac{1}{2} v_m B_m l (1 + \cos 2\omega t) \tag{10.61}$$

(assuming that the flexible leads contribute nothing).

Now we calculate the contribution due to the variation of **B**, assuming the wire or the rod to be 'frozen' at the position shown, so that:

$$\iint \left(\frac{\partial \mathbf{B}}{\partial t}\right) \cdot d\mathbf{S} = \int_{-a}^{z} (-\mathbf{i}_z \omega B_m \sin \omega t) \cdot (\mathbf{i}_z l dz)$$

$$= -\omega B_m l(z + a) \sin \omega t \tag{10.62}$$

Now from Eq. (10.57), $|\mathbf{v}'| = \dfrac{dz}{dt} = v_m \cos \omega t$

$$\therefore \qquad z = +\left(\frac{v_m}{\omega}\right) \sin \omega t$$

Hence, substituting for z in Eq. (10.62), we get

$$\iint\left(\frac{\partial \mathbf{B}}{\partial t}\right)\cdot d\mathbf{S} = -\omega B_{\mathrm{m}} la \sin \omega t - \frac{1}{2} B_{\mathrm{m}} l \mathrm{v}_{\mathrm{m}} (1 - \cos 2\omega t) \tag{10.63}$$

$\therefore$ Substituting from Eqs. (10.61) and (10.63) in Eq. (10.60), the induced emf is

$$\mathcal{E} = \oint_C \mathbf{E} \cdot d\mathbf{l} = B_{\mathrm{m}} \omega l a \sin \omega t - B_{\mathrm{m}} l \mathrm{v}_{\mathrm{m}} \cos 2\omega t \tag{10.64}$$

Thus, the output of the device contains a fundamental frequency term along with a second harmonic term, and ***no*** direct current.

10.7.4 A Rectangular Copper Plate Moving with Constant Velocity v through a Uniform Magnetic Field

As shown in Figure 10.27, the copper plate moves with a constant velocity **v** through a uniform magnetic field **B**. An electrostatic voltmeter is connected to the plate through the sliding contacts.

The force per unit charge, $\mathbf{F} = \mathbf{E} + \mathbf{v} \times \mathbf{B}$.

In this case, $\mathbf{v} = \mathbf{i}_z \mathrm{v}$, and $\mathbf{B} = \mathbf{i}_y B$.

$\therefore$
$$\mathbf{F} = (\mathbf{i}_z \times \mathbf{i}_y)\mathrm{v}B = -\mathbf{i}_x \mathrm{v}B \tag{10.65}$$

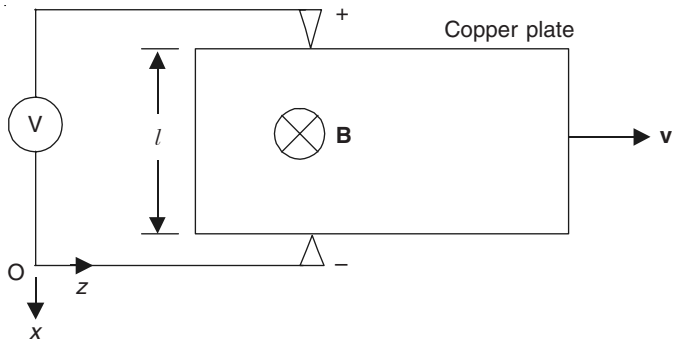


Figure 10.27 A rectangular plate type unipolar generator.

This causes the electrons to flow in the x-direction, with the positively charged region near the top, and the negatively charged region near the bottom. The equilibrium is reached when the electrostatic forces equal the motional force. The voltmeter reading is, then = Blv. In this case, there is no transformer emf.

10.7.5 An Expanding Rectangular Loop in a Time-varying Uniform Magnetic Field

A rectangular loop as shown in Figure 10.28, has a sliding conductor moving with a constant velocity **v**. The flux density **B** is perpendicular to the loop, and is uniform spatially in the whole region. The magnitude of the flux density, however, varies everywhere harmonically with time as $B = B_{\mathrm{m}} \cos \omega t$. In this problem,

$$\mathbf{B} = \mathbf{i}_z B_{\mathrm{m}} \cos \omega t \qquad \text{and} \qquad \mathbf{v} = \mathbf{i}_x \mathrm{v}$$

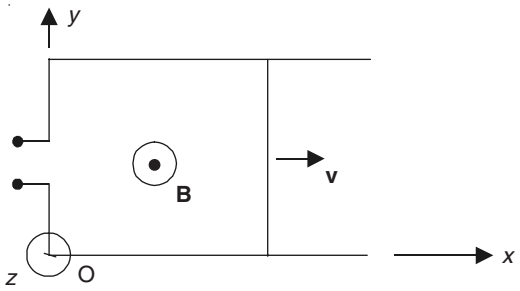


Figure 10.28 An expanding loop in a time-varying magnetic field.

$$\therefore \qquad E_{12} = -\iint_S \left(\frac{\partial \mathbf{B}}{\partial t}\right) \cdot d\mathbf{S} + \int_C (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{l}$$

Transformer emf

$$\frac{\partial \mathbf{B}}{\partial t} = -\mathbf{i}_z \omega B_{\mathrm{m}} \sin \omega t \qquad \text{and} \qquad d\mathbf{S} = \mathbf{i}_z \, dx\, dy$$

$$\therefore \qquad \left(\frac{\partial \mathbf{B}}{\partial t}\right) \cdot d\mathbf{S} = -\omega B_{\mathrm{m}} \sin \omega t \, dx\, dy$$

$$\therefore \qquad -\iint_S \left(\frac{\partial \mathbf{B}}{\partial t}\right) \cdot d\mathbf{S} = \omega B_{\mathrm{m}} \sin \omega t \int_{y=0}^{y=l} dy \int_{x=0}^{x=\mathrm{v}t} dx = +\omega l \mathrm{v} t B_{\mathrm{m}} \sin \omega t \tag{10.66}$$

Motional emf

$(\mathbf{v} \times \mathbf{B}) = -\mathbf{i}_y \mathrm{v} B_{\mathrm{m}} \cos \omega t$, the only contribution is from the sliding conductor, i.e. $d\mathbf{l} = \mathbf{i}_y dy$, and

$$\therefore \qquad (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{l} = -\mathrm{v} B_{\mathrm{m}} \cos \omega t \, dy$$

$$\oint_C (\mathbf{v} \times \mathbf{B}) \cdot \mathrm{d}\mathbf{l} = -\mathrm{v} B_{\mathrm{m}} \cos \omega t \int_{y=0}^{l} dy = -\mathrm{v} B_{\mathrm{m}} l \cos \omega t \tag{10.67}$$

$$\therefore \qquad E_{12} = \mathrm{v} B_{\mathrm{m}} l \, (\omega t \sin \omega t - \cos \omega t) \tag{10.68}$$

In this problem, if the velocity $\mathbf{v}$ is not a constant quantity, but varies exponentially with time which is given by:
$\mathbf{v} = \mathbf{i}_x f e^{gt}$, where the time is measured as zero at the instant the moving conductor leaves the y-axis of Figure 10.28. In this case:

$$\mathbf{v} = \mathbf{i}_x f e^{gt} = \frac{dx}{dt}$$

or

$$x = \int_{t=0}^{t} f e^{gt} dt = \left(\frac{f}{g}\right) e^{gt} \Big|_0^t = \left(\frac{f}{g}\right)(e^{gt} - 1)$$

and

$$\mathbf{B} = \mathbf{i}_z B_{\mathrm{m}} \cos \omega t$$

Transformer emf

$$\frac{\partial \mathbf{B}}{\partial t} = -\mathbf{i}_z \omega B_m \sin \omega t \qquad \text{and} \qquad d\mathbf{S} = \mathbf{i}_z dx\, dy$$

$$\therefore \quad -\iint_S \left(\frac{\partial \mathbf{B}}{\partial t}\right) \cdot d\mathbf{S} = +\omega B_m \sin \omega t \int_{y=0}^{y=l} dy \int_{x=0}^{x=(e^{gt}-1)f/g} dx$$

$$= \omega l \left(\frac{f}{g}\right)(e^{gt} - 1)\, B_m \sin \omega t \tag{10.69}$$

Motional emf

$$(\mathbf{v} \times \mathbf{B}) = -\mathbf{i}_y f e^{gt} B_m \cos \omega t \qquad \text{and} \qquad d\mathbf{l} = \mathbf{i}_y dy$$

$$\therefore \quad (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{l} = f e^{gt} B_m \cos \omega t\, dy$$

$$\oint_C (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{l} = -f e^{gt} B_m \cos \omega t \int_{y=0}^{y=l} dy = -f l B_m e^{gt} \cos \omega t \tag{10.70}$$

$$\therefore \quad E_{12} = f l B_m \left[\frac{\omega}{g}(e^{gt} - 1) \sin \omega t - e^{gt} \cos \omega t\right] \tag{10.71}$$

10.7.6 A Rectangular Loop Rotating in a Uniform Magnetic Field (Basis for an AC Generator)

We now consider a rotating rectangular loop in a steady magnetic field as shown in Figure 10.29. The loop rotates with an uniform angular velocity ω radians/second. This

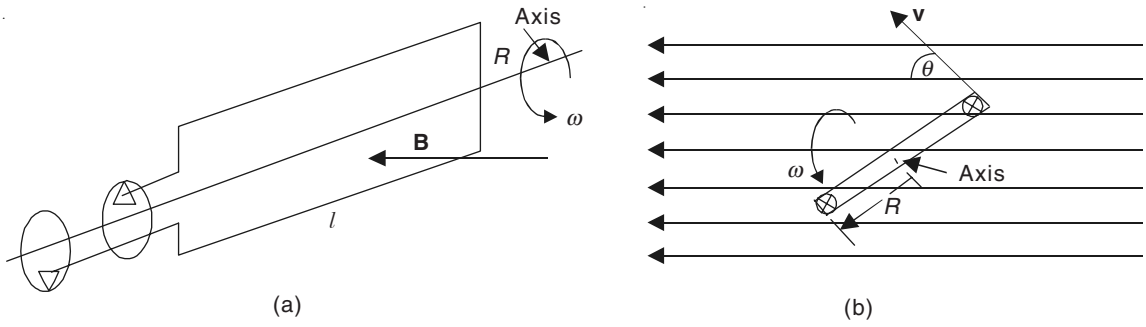


Figure 10.29 A simple AC generator: (a) perspective view and (b) cross-section perpendicular to the axis.

arrangement is essentially a simple alternating current generator, the induced emf appearing at the terminals connected to the slip-rings. If the radius of the loop is R (or width is $2R$) and its axial length l, we can evaluate the total induced emf in the loop as follows.

This is a case of motion only, and hence the total induced emf will be given by

$$\mathcal{E} = \oint (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{l} = 2vBl \sin \theta$$

Since

$$\theta = \omega t,\ \mathcal{E} = 2\omega RlB \sin \omega t \tag{10.72}$$

The factor 2 is due to the two conductors of length l moving through the **B** field and the emfs in both supporting each other. Also, $2Rl = A$, the area enclosed by the loop, and so:

$$\mathcal{E} = \omega BA \sin \omega t \tag{10.73}$$

Since $(\partial \mathbf{B}/\partial t) = 0$, there is no transformer emf, and the above equation gives the total induced emf.

Let us next consider the same loop rotating in an uniform magnetic field which is now varying harmonically with time at the same angular velocity as that of the rotation of the coil, i.e.

$$\mathbf{B} = \mathbf{B}_{\mathrm{m}} \sin \omega t$$

and hence, when $t = 0$, $\mathbf{B} = 0$ and $\theta = 0$.

The induced emf will now have both the components, i.e. the motional as well as the transformer. We first calculate the motional emf, i.e.

$$\mathcal{E}_{\text{motional}} = \oint (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{l} = 2\omega RlB_{\mathrm{m}} \sin^2 \omega t$$

$$= \omega RlB_{\mathrm{m}} - \omega RlB_{\mathrm{m}} \cos 2\omega t \tag{10.74}$$

The derivation is similar as for Eq. (10.72). And, next, due to the time-varying **B**,

$$\mathcal{E}_{\text{transformer}} = -\iint \left(\frac{\partial \mathbf{B}}{\partial t}\right) \cdot d\mathbf{S} = -2\omega RlB_{\mathrm{m}} \cos^2 \omega t$$

$$= -\omega RlB_{\mathrm{m}} - \omega RlB_{\mathrm{m}} \cos 2\omega t \tag{10.75}$$

$\therefore$ The resultant induced emf is

$$\mathcal{E} = \mathcal{E}_{\text{motional}} + \mathcal{E}_{\text{transformer}} = -2\omega RlB_{\mathrm{m}} \cos 2\omega t \tag{10.76}$$

Thus the resultant induced emf has a frequency which is twice that of the rotation of the coil or of the magnetic field. Though each component of the induced emf shows the presence of a time-independent component in it, the resultant emf has only the time-varying component, because the time-independent parts being of opposite polarity get cancelled out.

10.7.7 Faraday Disc or a Homopolar Generator (A Problem with Relative Motion between the Line of the Circuit and the Material Itself)

This is the second problem in which the Neumann's law seems to give the wrong result, the first one being in Section 10.7.2 in which case also, the apparent flux-linkage in the circuit appeared to be zero at all time . In the present problem, we have a circular conducting disc of conductivity σ rotating at a constant angular velocity of ω radians per second in an uniform steady magnetic field **B** (at right angles to the plane of the disc). The magnetic field can be generated by an N-turn coil wound on a surrounding magnetic circuit or by the poles of a permanent magnet.

When the voltmeter is connected across the device as shown in Figure 10.30(a), it will be seen that the flux-linkage across the circuit is permanently zero, and hence this would lead to the apparent conclusion that the induced emf in the circuit is zero.

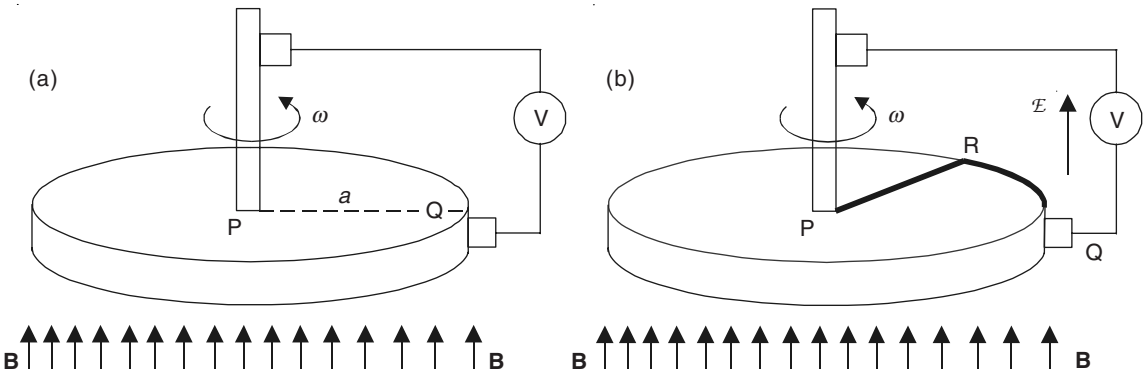


Figure 10.30 A Faraday disc.

A rigorous analysis of the problem would be discussed later when we consider the explanation of the electromagnetic phenomena as seen by an observer in relative motion, in Chapter 20 of this book. At that time, we shall discuss those aspects of electromagnetic induction, which we cannot analyze completely until we have written down the Maxwell's equations and the Lorentz force equation on the basis of the special relativity. At this stage, we offer the explanation given by Prof. G.W. Carter in his book , to show how the correct behaviour of the Faraday's disc can be understood and explained without going into the formal mathematics of special relativity. The explanation is as follows:

> *The equation* $\mathcal{E} = -(d\Phi/dt)$ *always gives the induced emf correctly, provided the flux-linkage is evaluated for a circuit so chosen that at no point are the particles of the material moving across it.*

Thus the emf generated by the Faraday's disc can be evaluated by the circuit shown in Figure 10.30(b). From the rim contact point Q, the circuit cannot proceed directly straight to the shaft P, but must follow a marked radius PR by which it makes its way to the shaft. Because this radius is fixed to the material of the disc, the flux-linkage of the sectoral area PRQ is continually increasing. Hence the induced emf is given by

$$\mathcal{E} = -\omega B \int_0^a r dr = \left(\frac{1}{2}\right)\omega B a^2$$

The direction of $\mathcal{E}$ is shown by the arrow.

The same method can be used for calculating the induced emf due to the rotating cylindrical bar magnet. The complete behaviour of the Faraday's disc and the Homopolar machine will be discussed in Chapter 20. However before closing this chapter, we shall tabulate (Table 10.1) the behaviour and the operating principles of different types of generators and motors, based on our study in the present chapter and Chapter 20.

Table 10.1 Common generators and motors

Type	*Rotor*	*Stator*	*Field production*	*Current distribution*	*Generator*	*Motor*
Homopolar	Metal Disc	Permanent magnet or pole pieces and field winding	Permanent magnet or separately excited	Slip rings	Faradays disc Lorentz disc Homopolar generator	Barlow's wheel
Heteropolar	Drum armature with multiple windings	Pole pieces and field winding	Self excited (a) Series wound (b) Shunt or compound wound separately excited	Commutator	DC generators	DC motors Small AC motors
Heteropolar	Drum armature	Pole piece and field windings	Separately excited	Slip rings	Small AC generators	AC synchronous motors
Heteropolar	Pole piece field windings	Stator windings one set per phase	Separately excited	Slip rings	Large AC generators	AC synchronous motors
Induction	Squirrel cage or conductors	Stator windings one set per phase				AC induction motors

PROBLEMS

10.1 Sketch the current waveform when a direct voltage is applied to a pure inductance. What limits the current and what determines the initial rate of rise of current, in a practical coil?

10.2 A single-turn circular coil, having a diameter much greater than the thickness of the wire, has a resistance of 0.012 Ω. When the coil carries a 50 Hz current of 5 amps, the voltage across the coil is 65 mV. Estimate the open-circuit induced emf in an exactly similar coil taped to the first.

Ans.: 25 mV

10.3 A solenoid is wound on a long former of square cross-section and containing no magnetic material. It is bent round into a toroid of internal and external radii a and b respectively. A straight thin cable of infinite length passes along the axis of the toroid at right angles to its plane. Show that the mutual inductance between the cable and the toroid is

$$M = \mu_0 n \frac{b^2 - a^2}{2} \ln\left(\frac{b}{a}\right) \text{ henry}$$

where n is the mean number of turns per metre on the solenoid.

10.4 Calculate the inductance of a 500-turn coil wound on a toroidal core having an outer diameter of 15 cm, inner diameter of 10 cm, a squae cross-section and a relative permeability of 100. What error will be introduced by assuming that the magnetic flux density was equal to the flux density at the mean diameter multiplied by the area.

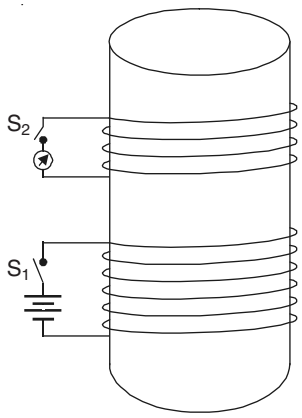
Ans.: $L = 0.4$ mH
% error 1.2%

10.5 A steel bar with circular cross-section has two windings on it as shown in the adjoining figure.

At first the switch S_1 in the exciting winding is permanently closed, and the switch S_2 in the second winding is opened and closed periodically.

In the second experiment, the switch S_2 is permanently closed, and the switch S_1 in the exciting winding is opened and closed periodically.

How would the galvanometer behave in both these experiments? Give reasons for your conclusions.

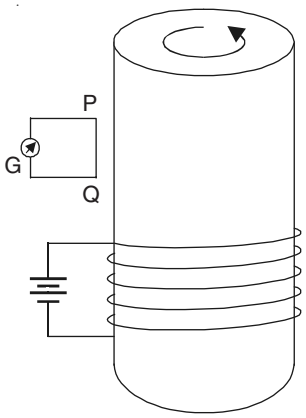


10.6 A steel bar of circular cross-section is wound with an exciting winding as shown in the adjoining figure.

At first the steel bar is rotated at a certain speed, keeping everything else stationary.

In the second experiment, only the circuit PQG is made to revolve round the bar such that PQ is always parallel to the axis of the bar and the plane of the circuit is always in a fixed r–z plane. Everything else is now stationary.

How would the galvanometer in the circuit PQG behave in these two experiments? Give reasons for your conclusions.



11 Forces and Energy in the Static and Quasi-Static Magnetic Fields

11.1 ENERGY RELATIONS IN QUASI-STATIONARY ELECTROMAGNETIC FIELDS

11.1.1 Introduction

It should be understood that to establish any static system it is necessary to have a transient, non-stationary process until the final, static state is attained. This is true even in the case of the electrostatic fields where the transient process implies a 'non-electrostatic' process of bringing the charges to their desired positions or on to the conducting bodies. In the case of the static magnetic fields, this transient process starts as soon as the currents are switched on, or (theoretically) when the current systems are being brought from the infinity, and this (transient) process ends when all the currents of the system can be considered as steady.

During the transient process, the charging currents would induce certain electromotive forces in the relevant contours, though at the end of the process no induced emf's would remain. In order to calculate the energy of the system, i.e. the energy required to establish the desired current distributions, the existence of these induced emf's must be taken into account. It is for this reason that before analyzing the energy relations in a static magnetic field we had to learn and understand the phenomenon of electromagnetic induction. Now we shall consider, more generally the quasi-stationary systems and then the relations for the static magnetic systems would be deduced as a special case of the more general system.

11.1.2 The Energy Required to Establish a Magnetic Field

Let us assume an arbitrary quasi-stationary current system produced by a time-varying electric field $\mathbf{E}_i$. The current density $\mathbf{J}$ of the time-varying currents produced by this field is

$$\mathbf{J} = \sigma(\mathbf{E}_i + \mathbf{E}_{tot}) = \sigma\left(\mathbf{E}_i - \frac{\partial \mathbf{A}}{\partial t} - \text{grad } V\right) \tag{11.1}$$

where $\mathbf{E}_{tot}$ is due to the quasi-stationary charges on the conductor surfaces (the term $-\text{grad } V$) and to the varying currents (i.e. the accelerated charges) inside the conductors (the term $-\partial\mathbf{A}/\partial t$). To

obtain the energy relations, we first determine the power required to maintain the time-varying currents and hence the corresponding magnetic fields. So we multiply Eq. (11.1) by $\mathbf{J}dv/\sigma$, where dv is the volume element in which the current density is $\mathbf{J}$ and the conductivity σ. Integrating the product over the whole volume v, we get

$$\iiint_v \mathbf{E}_i \cdot \mathbf{J}\, dv = \iiint_v \left(\frac{J^2}{\sigma}\right) dv + \iiint_v \mathbf{J} \cdot \left(\frac{\partial \mathbf{A}}{\partial t}\right) dv + \iiint_v (\mathbf{J} \cdot \nabla V) dv \tag{11.2}$$

The left-hand side of this equation represents the power of the external forces acting on the free charges in the system, and the right-hand side terms describe how this power is being used up.

The first term on the right represents the power which is being transformed into heat in the whole system. This is known as the 'Joule's heat'. The third term is that part of the power of the external forces which is used against the forces of the electric field created by the quasi-stationary distribution of the charges. The second term represents that part of the power of the external forces which is used against the *forces of the quasi-stationary electric field* $(\partial \mathbf{A}/\partial t)$ *created by all the varying currents of the system at the instant considered.* It is this term which is associated with the energy required to establish some current distributions and hence the magnetic field. This we denote by P_m,

$$P_m = \iiint_v \mathbf{J} \cdot \left(\frac{\partial \mathbf{A}}{\partial t}\right) dv \tag{11.3}$$

This equation is valid generally, and the only restriction imposed on this relationship is that the field must be quasi-stationary. The region v may contain nonlinear, ferromagnetic materials, or even a part of P_m may be used for doing mechanical work in moving some bodies of the system by means of the magnetic forces. For linear stationary media, this equation can be simplified, as we can write, $\mathbf{J} = \mathbf{J}_0 f(t)$ where $\mathbf{J}_0$ is a function of position only. So the magnetic vector potential $\mathbf{A}$ also must be of similar form, i.e. $\mathbf{A} = \mathbf{A}_0 f(t)$. By substituting in Eq. (11.3), it reduces to

$$P_m = \left(\frac{\partial}{\partial t}\right) \iiint_v \left(\frac{1}{2}\right) \mathbf{J} \cdot \mathbf{A}\, dv \tag{11.4}$$

To obtain the energy, we multiply these equations by dt (a small time element) and integrate them with respect to time over the interval $t = 0$ to $t = t$, and we get

$$W_m = \int_0^t \iiint_v \mathbf{J} \cdot \left(\frac{\partial \mathbf{A}}{\partial t}\right) dt\, dv = \int_0^t \iiint_v \mathbf{J} \cdot d\mathbf{A}\, dv \tag{11.5}$$

and

$$W_m = \int_0^t dt \iiint_v \left(\frac{1}{2}\right) \mathbf{J} \cdot \mathbf{A}\, dv = \left(\frac{1}{2}\right) \iiint_v \mathbf{J} \cdot \mathbf{A}\, dv \tag{11.6}$$

Equation (11.5) is valid generally, and Eq. (11.6) is restricted to the linear stationary media.

11.2 POTENTIAL ENERGY OF A CIRCUIT IN A MAGNETIC FIELD

We now consider a specific circuit placed in a magnetic field. So we take a circular coil carrying a current I, placed in a magnetic field $\mathbf{B}$. This will be subject to a torque $\mathbf{T}$, given by

$$\mathbf{T} = I\mathbf{S} \times \mathbf{B} \tag{11.7}$$

where **S** is the vector area covered by the coil (Figure 11.1).

$$\therefore \qquad T = ISB \sin \theta$$

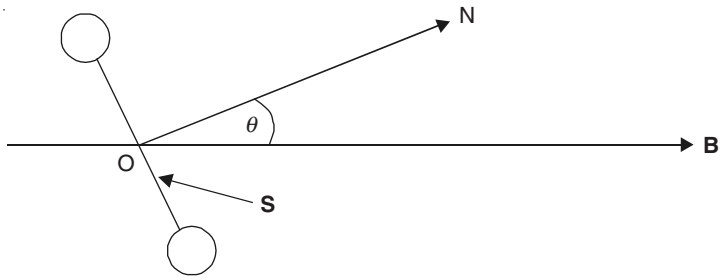


Figure 11.1 A circular coil in a magnetic field.

The torque exerted on the coil tends to turn it so that its normal aligns itself with the direction of **B** (so as to maximize its flux-linkage). If now we consider an extraneous torque twisting the coil away from this direction of alignment, then the work done by this extraneous torque against the generated torque **T** may be considered to be stored as the 'potential energy'. If the current loop is now allowed to turn back, then the stored energy is recoverable.

Let us now set an arbitrary datum line, i.e. let the potential energy be = 0, when $\theta = 0$.

$\therefore$ Potential energy U when $\theta = \theta$ (a finite value), is

$$= \int_0^{\theta} ISB \sin \theta \, d\theta = ISB(1 - \cos \theta) \tag{11.8}$$

where $B \cos \theta$ = the component of **B** in the direction of the coil-axis ON.

$\therefore$ $SB \cos \theta = \Phi$ = the flux linked by the coil; and

$SB = \Phi_0$ = the flux linked by the coil when its axis is aligned with the direction of **B**.

$$\therefore \qquad U = I\Phi_0 - I\Phi = U_0 - I\Phi \tag{11.9}$$

It should be noted that U_0 is a constant for the coil. Since, in general, we are concerned with the energy differences (in only two positions), in these subtractions U_0 will cancel out. So, without any loss of generality, we can drop out the constant U_0 from the expression for the potential energy, i.e.

$$U = -I\Phi \tag{11.10}$$

In fact, this amounts to saying that the potential energy is zero when $\theta = \pi/2$, that is when no flux is linked with the coil. In the next step, we now generalize it for any extensive circuit which can be considered to be made up of a mesh-work of small circular loops, all carrying the same current I. Hence, we get

$$\begin{aligned} U &= U_1 + U_2 + U_3 + U_4 + \cdots \\ &= -I(\Phi_1 + \Phi_2 + \Phi_3 + \Phi_4 + \cdots) \\ &= -I \times \text{Flux through the extensive circuit} \end{aligned} \tag{11.11}$$

11.2.1 Location of the Stored Energy of a Circuit

Let a circuit be so moved or deformed that the flux linked by it is reduced by an amount denoted by $\delta\Phi$.

Then, the increase in its potential energy, $\delta U = I\delta\Phi$.

Let the current be maintained at the constant value I, during this process of change, by some external source of power, and let the time interval required for this flux change be δt.

Then, by the Neumann's law, the emf induced in the circuit during this time interval is

$$\mathcal{E} = \frac{\delta\Phi}{\delta t}$$

and, since, $\delta\Phi$ denotes the decrease in the flux, the direction of $\mathcal{E}$ would be such as to support the flow of the current.

$\therefore$ The energy drawn from the battery is reduced by an amount

$$= \mathcal{E}I\delta t = I\delta\Phi = \delta U$$

the battery is now storing the energy. On the other hand, if the circuit is so deformed as to increase the flux-linkage, the energy from the battery is drawn out. Thus it is seen that the equation for the induced emf and the potential energy of the circuit are linked with each other, by the way of the 'principle of conservation of energy'.

11.3 FORCES ON A CURRENT-CARRYING CIRCUIT IN A MAGNETIC FIELD

So from the knowledge of the potential energy of a circuit, it is possible to calculate the force on a current-carrying circuit placed in a magnetic field.

Let us assume a deformable circuit carrying a current I placed in a magnetic field **B** (Figure 11.2), which makes an angle θ with the direction of the element δl. If the element is moved in the direction of the circuit [Figure 11.2(b)] or in the direction of **B** [Figure 11.2(c)], then in either case, there has been no change in the flux-linkage of the circuit, and hence no change in its potential energy, and so

the work done = 0

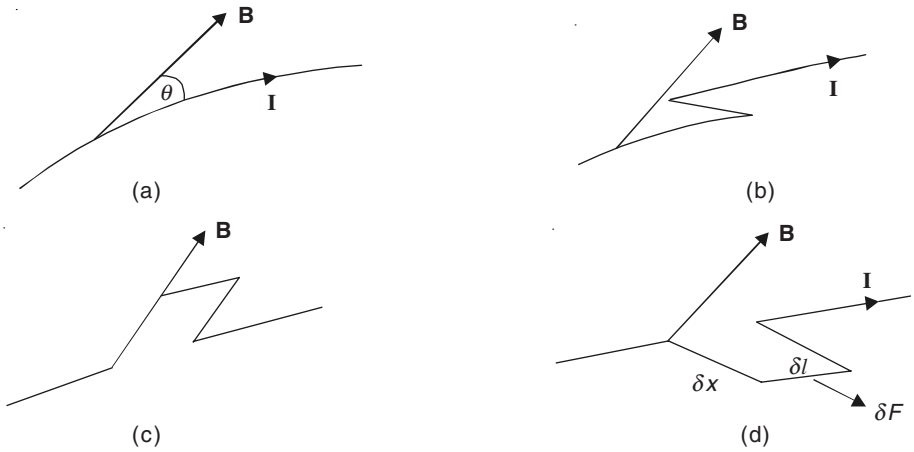


Figure 11.2 A deformable circuit in a magnetic field **B**.

$\therefore$ The electromagnetic force on the element δl must act normally to both **B** as well as the element δl of the circuit. If, now, under the influence of this force $\delta\mathbf{F}$, the element δl is allowed to move a distance δx, in the direction of $\delta\mathbf{F}$, then the extra flux linked by the circuit [Figure 11.2(d)] due to this deformation is

$$\delta\Phi = B \sin\theta\, \delta x\, \delta l$$

$\therefore$ Decrease in potential energy $= IB \sin\theta\, \delta x\, \delta l$

$$= \text{work done by } \delta\mathbf{F}$$

$$= \delta\mathbf{F} \cdot \delta\mathbf{x}$$

$\therefore$
$$\delta F = IB \sin\theta\, \delta l \tag{11.12}$$

and $\delta\mathbf{F}$ is perpendicular to both $\delta\mathbf{l}$ and **B** such that it tends to increase the flux linked by the circuit, (i.e. a rotation from the direction of I to the direction of **B** is right-handed about the direction of **F**).

If the wire has a finite cross-section A, so that the current density is **J**, then $I = JA$. Also, $A\delta l = \delta v$—a volume element in the conductor. Hence, Eq. (11.12) can also be written as

$$\delta F = JB \sin\theta\, \delta v \tag{11.13}$$

11.4 FORCE ON A MOVING CHARGE IN A MAGNETIC FIELD

While studying the electromagnetic force on the current-carrying wires, J.C. Maxwell had commented that the mechanical force impelling a current-carrying conductor across the lines of magnetic force, acts on the conductor and not on the current in the conductor. This conclusion was based on the experimental observation that when an electric current flows across a magnetic field, there is a force pushing the conductor sideways, but "has *no* tendency for the line of flow of the current to move within the conductor". This latter part of the statement has been shown to be strictly incorrect as we have already seen in Section 9.3.1. This effect (i.e. the Hall effect) is a consequence of the Lorentz force component acting upon the free charges in the current-carrying media. This effect can be well observed in the beam of electrons in a cathode-ray tube. It should be understood that the force is primarily on the electrons. But when the electrons are moving in a conductor under the influence of the force, their (i.e. the electrons) motion gets halted by the collisions with the lattice structure of the conducting material, as a result of which the force gets transferred from the charges to the conductor which carries them.

We now consider a tubular stream of charges Q moving with an average velocity v. Let there be N charges per unit volume of the stream and its cross-section be A. In a time-interval δt, the charges in a cylinder of length $\text{v}\delta t$ would cross a section of the tube at P, and the total charge in the cylinder is $NQ\text{v}A\delta t$ (Figure 11.3).

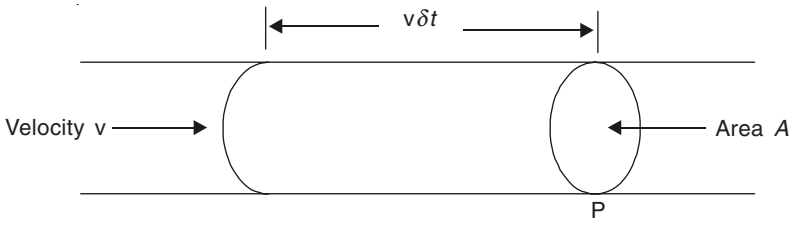


Figure 11.3 A stream of moving charges.

$\therefore$ The current in the stream = $NQvA$, and

$$\text{the current density } J = NQv \tag{11.14}$$

$\therefore$ The force δF on a volume element δv of the stream is

$$\delta F = (NQvB \sin \theta)\delta v$$

But the number of the charges in this volume element is $N\delta v$.

$\therefore$ The force on each charge is

$$F_m = QvB \sin \theta \tag{11.15}$$

The direction of this force, for positive Q, is at right angles to the plane containing the vectors **v** and **B** and its sense is such that a rotation from the direction of **v** to the direction of **B** is right-handed about the direction of $\mathbf{F}_m$, i.e.

$$\mathbf{F}_m = \mathbf{v} \times \mathbf{B} \tag{11.16}$$

When the charges are electrons, the direction of the force is opposite to that mentioned above. This formula has been well checked experimentally.

11.5 CALCULATION OF FORCES ON CIRCUITS (BASED ON THE POTENTIAL ENERGY)

Using the knowledge of the potential energy of the system, the forces on many forms of circuits can be calculated. This can be done either from the energy of the whole circuit, or by adding up the contributions from each element of the circuit. If a circuit is bodily moved in a direction by a distance δx, and the force in that direction is F_x, then

$$\text{the decrease in the potential energy} = F_x\, \delta x\, (= -\, \delta U)$$

$\therefore$
$$F_x = -\frac{\partial U}{\partial x} = I\left(\frac{\partial \Phi}{\partial x}\right) \tag{11.17}$$

Similarly, if a circuit rotates about an axis by an angle $\delta\theta$, then the torque about the axis is

$$T_\theta = -\frac{\partial U}{\partial \theta} = I\left(\frac{\partial \Phi}{\partial \theta}\right) \tag{11.18}$$

We shall now consider some examples of forces on some typical circuits.

11.5.1 Forces between Stationary Current Systems (The Ampere–Laplace Law or the Ampere's Law of Forces)

We consider the two current-carrying conductors with steady currents, the charges moving inside which act on each other by the forces given by Eq. (7.1). Thus the resultant force on the two conductors is equal to the sum of the forces between all the pairs of the elemental charges. To determine the macroscopic forces between the two conductors, instead of adding up all these elemental forces together, we consider the average velocity and the density of the charges, and then the total force is taken as an infinite sum of the elemental forces acting between the pairs of physically small volumes with the moving charges inside them.

So let us consider two elemental volumes δv_1 and δv_2 inside a stationary or quasi-stationary current field (Figure 11.4). These elements can belong to the same current-carrying conductor or may be of two different conductors, between which the force is to be determined. Now, let the charge of the free charge carriers be Q, their average velocities in δv_1 and δv_2 be $\mathbf{v}_1$ and $\mathbf{v}_2$, respectively, and there are N_1 and N_2 charges per unit volume of the two systems respectively. If r is the distance between δv_1 and δv_2, then the magnetic force on all the moving charges in δv_2, due to all the moving charges in δv_1 is given by

$$\delta \mathbf{F}_{m12} = \left(\frac{\mu_0}{4\pi}\right)\left(\frac{N_2 Q\mathbf{v}_2 \times (N_1 Q\mathbf{v}_1 \times \mathbf{u}_{12})}{r^2}\right)\delta v_1 \delta v_2$$

$$= \left(\frac{\mu_0}{4\pi}\right)\left[\frac{\mathbf{J}_2 \times (\mathbf{J}_1 \times \mathbf{u}_{12})}{r^2}\right]\delta v_1 \delta v_2 \tag{11.19}$$

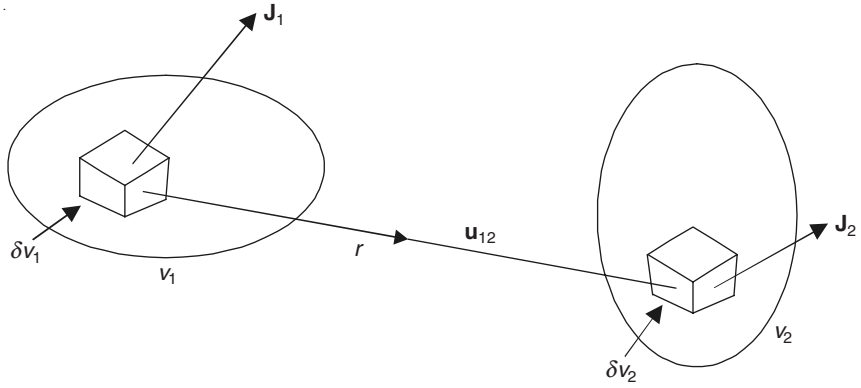


Figure 11.4 Two current-carrying regions.

Let δv_1 and δv_2 be the volume elements of the macroscopic current-carrying volumes v_1 and v_2 respectively. Hence the total force by which the currents in v_1 act upon the currents in v_2 is given by

$$\mathbf{F}_{m12} = \left(\frac{\mu_0}{4\pi}\right)\iiint_{v_2}\iiint_{v_1}\left[\frac{\mathbf{J}_2 \times (\mathbf{J}_1 \times \mathbf{u}_{12})}{r^2}\right]dv_1\, dv_2$$

$$= \left(\frac{\mu_0}{4\pi}\right)\iiint_{v_2}\left(\frac{\mathbf{J}_2 \times \mathbf{B}_1}{r^2}\right)dv_2 \tag{11.20}$$

by Eq. (7.16), according to which the integral over v_1 is the magnetic flux density vector $\mathbf{B}_1$ produced by the currents in the volume v_1, at the points in v_2. This is the most general formula for the magnetic force acting between the two current-carrying regions.

Now we consider the practically important case of the two quasi-filamentary current-carrying contours with the currents I_1 and I_2. Then we can write as

$$\mathbf{J}_1 \delta v_1 = \mathbf{J}_1(\Delta \mathbf{S}_1 \cdot \delta \mathbf{l}_1) = I_1 \delta \mathbf{l}_1 \qquad \text{and} \qquad \mathbf{J}_2 \delta v_2 = \mathbf{J}_2(\Delta \mathbf{S}_2 \cdot \delta \mathbf{l}_2) = I_2 \delta \mathbf{l}_2$$

Hence for the two closed contours C_1 and C_2 (Figure 11.5), Eq. (11.20) becomes

$$\mathbf{F}_{m12} = \left(\frac{\mu_0 I_1 I_2}{4\pi}\right) \oint_{C_1} \oint_{C_2} \left[\frac{d\mathbf{l}_2 \times (d\mathbf{l}_1 \times \mathbf{u}_{12})}{r^2}\right] \tag{11.21}$$

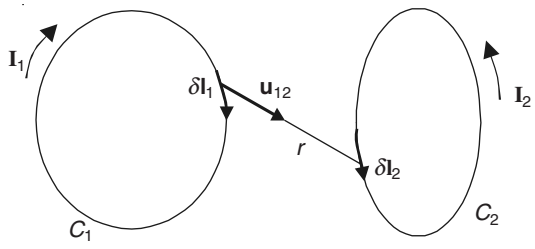


Figure 11.5 Two closed current filaments.

This expression for the force between the two filamentary current-carrying contours is referred to as the Ampere–Laplace law or sometimes as the Ampere's force law.

Reminding ourselves that the 'Biot–Savart's law' gives the magnetic flux density created by a current in a contour as [see Eq. (7.17)]

$$\mathbf{B}_{12} = \left(\frac{\mu_0 I_1}{4\pi}\right) \oint_{C_1} \frac{d\mathbf{l}_1 \times \mathbf{u}_{12}}{r_2^2}$$

at the point where the element dl_2 of the contour C_2 is located. Hence we can rewrite the force equation (11.21) in the general form (dropping all the subscripts)

$$\mathbf{F}_m = \oint_C I\, d\mathbf{l} \times \mathbf{B} \tag{11.22}$$

Hence if a current element ($I\, dl$) is situated in a magnetic field of flux density **B**, the elemental force on the current element is given by

$$\delta\mathbf{F}_m = I\, d\mathbf{l} \times \mathbf{B} \tag{11.23}$$

As an application of this method, we consider the force between two infinitely long, parallel conductors, carrying the same current I, but in opposite directions [parallel go-and-return circuit (Figure 11.6). The distance between the conductors is d. The magnetic flux density created by the current I in the conductor (1), at points of the conductor (2) is

$$B = \frac{\mu_0 I}{2\pi d}$$

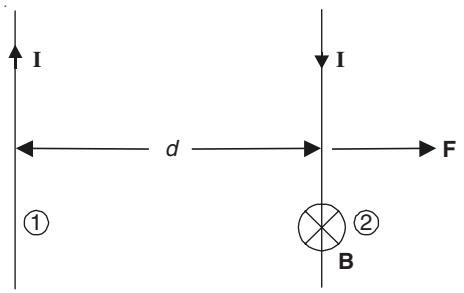


Figure 11.6 Parallel go-and-return circuit.

Its direction would be normal to the plane of the paper and into it.

∴ The force **F** per unit length, and acting outwards is

$$F = IB = \frac{\mu_0 I^2}{2\pi d} \tag{11.24}$$

This equation will also hold if the wires have finite cross-section provided the current is distributed uniformly across the conductor cross-section.

If the parallel conductors carried the currents in the same direction, then the only difference would be that the force between the two conductors would be attractive instead of repelling.

11.5.1.1 Time-varying currents in two parallel conductors

So far we have discussed only the effects of the time-independent direct currents, but when the two parallel conductors carrying alternating currents (at low power frequencies, i.e. at 50 Hz or 60 Hz, say) i.e. $I \sin \omega t$ in each, then the force as given by Eq. (11.24) would get modified to

$$F \text{ (per unit length)} = \left(\frac{\mu_0 I^2}{2\pi d}\right) \sin^2 \omega t$$

$$= \left(\frac{\mu_0 I^2}{2\pi d}\right)\left(\frac{1 - \cos 2\omega t}{2}\right) \tag{11.24a}$$

i.e. apart from a time-independent component of the force which is unidirectional in nature and has the magnitude $[\mu_0 I^2/(4\pi d)]$, there is also a vibratory (or oscillatory) component which alternates at double the frequency of the alternating currents producing it. This force is quite well-known in the machine design problems, in particular, for the stator windings of the turbo-alternators, where the stator slots have two parallel conductors in each, which carry alternating currents of the same phase or of different phases [Figure 11.6(a)]. These vibratory forces are known as the 'bar bouncing forces', and have to be taken into account for designing the slot wedges and the spring linings between the conductors and the slot walls.

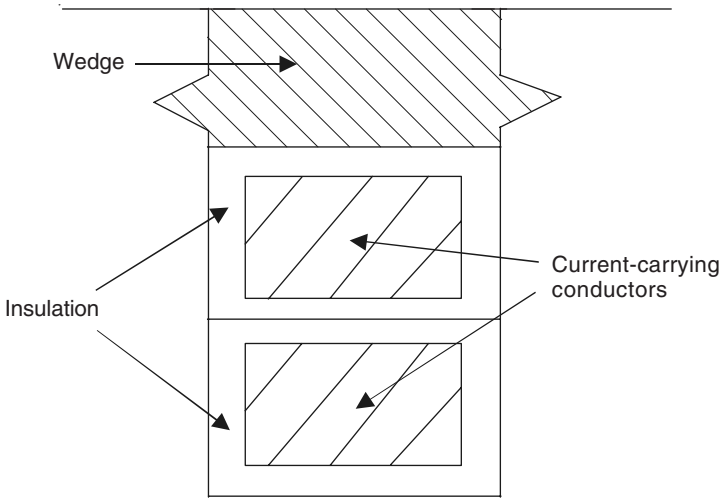


Figure 11.6(a) Stator slot of a turbogenerator.

These vibratory forces are somewhat more complicated than as indicated above because in these machines, the conductors in some of the slots carry currents of different phases. In such slots, the vibratory component of the force would be of the form [sin ωt sin ($\omega t \pm 2\pi/3$)] instead of $\sin^2 \omega t$. So now, the vibratory part of the force would now contain both the variations, i.e. cos $2\omega t$ as well as sin $2\omega t$, and the peak value of the force required for the design purposes has to be suitably modified. It should also be noted that under unbalanced load conditions or some of the fault conditions, the conductor current can contain a direct current component (or a zero sequence component), which would produce a vibratory force of the frequency ω in addition to the double frequency force.

11.5.2 Torque on a Circular Coil Placed in a Uniform Magnetic Field

This is a problem which arises with many measuring instruments. Let us consider a circular coil of radius a, the total ampere-turns I, and the axis of the coil (ON) making an angle θ with the direction of the magnetic field **B** (Figure 11.7).

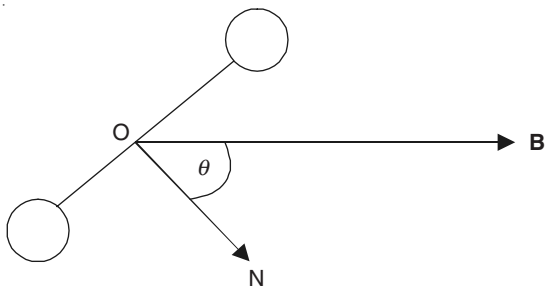


Figure 11.7 A circular coil in a magnetic field.

$\therefore$ The flux through the coil = $\pi a^2 B \cos \theta$

$\therefore$ From Eq. (11.18) , the torque is given by

$$T_\theta = I \frac{\partial}{\partial \theta} (\pi a^2 B \cos \theta) = - I \pi a^2 B \sin \theta \qquad (11.25)$$

The negative sign indicates that the torque tends to reduce the angle θ. If, for πa^2 (the area of the coil), we write S, the formula becomes

$$T_\theta = - ISB \sin \theta = - \mathbf{I} \cdot |\mathbf{S} \times \mathbf{B}|$$

This is the same expression as for the torque of small elemental coils, i.e. Eq. (8.6).

$\therefore$ Any coil in a uniform magnetic field **B** may be treated as a 'small' coil and dealt with as though it were a current loop.

11.5.3 Forces between the Turns of a Coil

In this problem, the turns of the coil are considered not as a helix, but as a set of coaxial rings of radius a and the current I in the same direction [Figure 11.8(a)].

On a particular ring, the flux density due to the other rings = **B**, inclined to the axis of the coil at an angle α.

$\therefore$ The axial component of the magnetic field = $B \cos \alpha$

and the radial component = $B \sin \alpha$

We consider the effects of the axial component of **B** first. This component reacts with the current I in the turn to produce a uniform outward pressure = $BI \cos \alpha$ per unit length.

This sets up a tension in the wire. This tension can be calculated by considering the equilibrium of half the turn under the influence of these pressures, and the tensions F at the ends of the semi-circle as shown in Figure 11.8(b).

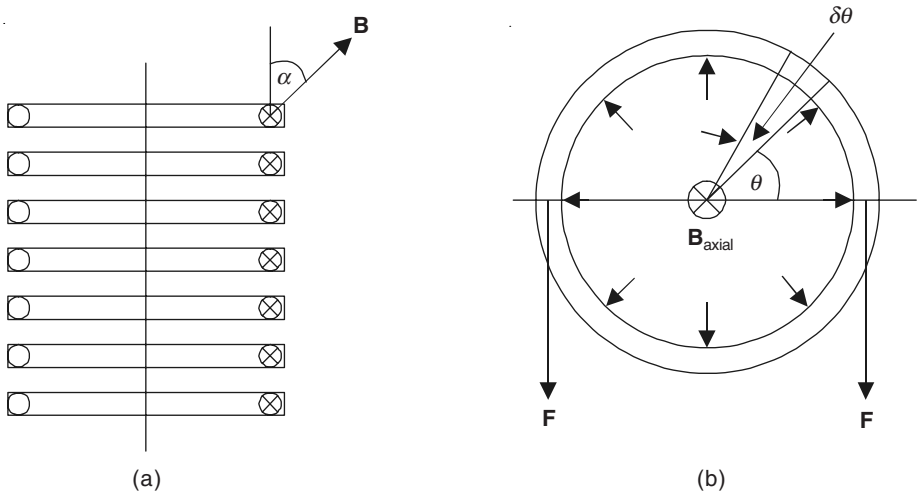


Figure 11.8 Forces between the turns of a multiturn coil.

The thrust on an arc $\delta\theta$ is = $(BI \cos \alpha)(a\ \delta\theta)$.

The component of the thrust in the direction of **F**, is

$$= (BI \cos \alpha)(a\ \delta\theta \sin \theta)$$

We consider all such elements in the semi-circle, and hence:

$$2F = \int_0^{\pi} BIa \cos \alpha \sin \theta\, d\theta$$

$$\therefore \qquad F = BIa \cos \alpha \tag{11.26}$$

The $2F$ is due to F at both the ends of the semi-circle.

Next, the pressure produced by the radial component of the field = $BI \sin \alpha$ per unit length (Figure 11.9).

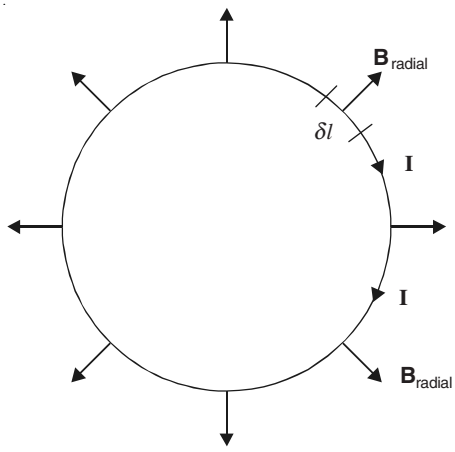


Figure 11.9 Effect of the radial component of **B** on the circular coil-turn.

$\therefore$ Force on each element = $BI \sin \alpha \, \delta l$, in the axial direction.

$\therefore$ The total axial force on the turn = $BI \sin \alpha \int_{\theta=0}^{\theta=2\pi} a \, d\theta$

$$= (BIa \sin \alpha)(2\pi) = 2\pi \, BIa \sin \alpha \tag{11.27}$$

As all the turns carry the current in the same direction, the direction of the axial force would be attractive between them, i.e. a compressive force between the turns. Furthermore, the flux lines near the ends of the coil are divergent, and so it will be seen that near the ends, the turns are subject to a larger component of the radial flux which would produce a larger axial force in the direction of the main body of the coil, i.e. the place where the field is the strongest.

Thus, when a power-system reactor (which is basically a large air-cored solenoid, mounted on a suitably structured insulating former and anchored to it) is damaged by the electromagnetic forces produced by the passage of excessively large currents (in the range of 10–20 kA), the turns of the reactor winding get crushed together, whilst each turn shows also the existence of a force which has tended to tear it apart. This is obvious from the photographs (see Plates 1–3).

The behaviour of the large transformer coils under the effect of such electromagnetic forces is somewhat different from that of the reactor coils discussed above. In a three-phase transformer, each core-limb carries both the h.v. and the l.v. coils which are mounted concentrically on the limb. If such a transformer is damaged during some short-circuit conditions, it seems that near the coil-ends there is some elongation of the outer end-turns of the outer coil which is the l.v. coil. This behaviour can be easily explained by considering the direction of the magnetic flux in the vicinity of these turns and also the direction of the currents in these turns. Since the h.v. and the l.v. coils carry currents in the opposite directions, and since near the coil-ends there would be a larger component of the radial flux, the end-turns would be subject to the repelling forces in the axial direction, tending (the l.v. turns) to move them away from the central coil-turns. This accounts for the elongation of the end-turns of the transformer coils.

11.5.4 The Motion of a Charged Particle in Magnetic and Electric Fields

11.5.4.1 The motion of the electric charges in a vacuum under the influence of the electric field

We consider a point charge Q of mass m, moving in a vacuum under the influence of an electric field **E**. The differential equation for the motion of the particle is

$$m \frac{d^2 \mathbf{r}}{dt^2} = Q\mathbf{E} \tag{11.28}$$

In the Cartesian coordinates, in the scalar form, this equation becomes

$$m \frac{d^2 x}{dt^2} = QE_x, \quad m \frac{d^2 y}{dt^2} = QE_y, \quad m \frac{d^2 z}{dt^2} = QE_z \tag{11.29}$$

Analysis of the motion of the charge implies solving the above equations for some specified initial conditions. Very often, when the final velocity of a charged particle accelerated by the electric field is required, we need the information regarding the potential difference between the initial and the final points of the particle path.

Work done by the electric field in moving the charge for a length δl is

$$= dW_e = Q\mathbf{E} \cdot \delta\mathbf{l}$$

This work is done to increase the velocity of the particle, i.e. increase its kinetic energy (K.E.).

Let the velocity at the initial point 'O' of the path be = v_0. Then the velocity v_1 at any point '1' of the path is

$$\left(\frac{1}{2}\right)mv_1^2 - \left(\frac{1}{2}\right)mv_0^2 = Q\int_0^1 \mathbf{E} \cdot d\mathbf{l} = Q(V_0 - V_1) \tag{11.30}$$

This, in fact, is the 'law of conservation of energy' in this case.

$$\therefore \qquad v_1 = \sqrt{v_0^2 + \frac{2Q(V_0 - V_1)}{m}} \tag{11.31}$$

When the initial velocity is zero, i.e. $v_0 = 0$, then

$$v_1 = \sqrt{\frac{2Q(V_0 - V_1)}{m}} \tag{11.32}$$

A practical application of this problem is the system of accelerating electrons in the electron gun of the cathode ray tubes.

11.5.4.2 The motion of a charged particle in a static magnetic field

Again we consider a charged particle of mass m and charge Q, moving in a vacuum in a magnetic field of flux density **B**. The magnetic force on the charge is given by the Lorentz force equation, in this case the electric field **E** being = 0.

$$\mathbf{F}_m = Q(\mathbf{v} \times \mathbf{B})$$

where **v** is the velocity of the particle. So the equation of motion of the particle is

$$m\frac{d\mathbf{v}}{dt} = Q(\mathbf{v} \times \mathbf{B}) \tag{11.33}$$

Since $(\mathbf{v} \times \mathbf{B})$ is perpendicular to the instantaneous velocity **v**, the kinetic energy of the particle and the magnitude of the velocity **v** are constant. In general, **v** would be evaluated by solving the three scalar differential equations of (11.33) subject to the specified initial conditions of the problem.

We first consider the case, where **B** is static and uniform. If the initial velocity is assumed to be parallel to **B**, then $\mathbf{F}_m = 0$ and the motion of the particle will not be affected by the magnetic field **B** and the particle will move with its initial velocity along the lines of **B**.

Next, we consider the case, when the initial velocity is perpendicular to **B**. Then, $|\mathbf{F}_m| = Qv B$. At each point of the path of the particle, the force on the charge (= $\mathbf{F}_m$) is at right angles to the motion and hence the velocity remains constant. Thus the particle experiences a constant acceleration, at every point of its motion, and the acceleration is also at right angles to the direction of the motion. Its path is therefore a circle with its radius R which is given by

$$\frac{mv^2}{R} = QvB, \quad \text{and} \quad \therefore \quad R = \frac{mv}{QB} \qquad \text{(see Figure 11.10)} \tag{11.34}$$

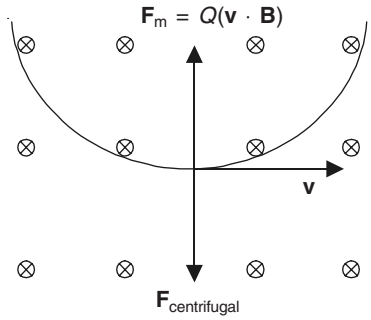


Figure 11.10 A charged particle moving at right angles to a uniform magnetic field.

The angular frequency of the particle, then, is

$$\omega_C = \frac{v}{R} = \frac{QB}{m}$$

and the true frequency is

$$f_C = \frac{\omega_C}{2\pi} = \frac{QB}{2\pi m}$$

This is known as the 'Cyclotron frequency' which depends only on the Q/m of the particle and the magnetic flux density. It is independent of the radius and the velocity (so long as $v << c$).

We now consider the next general case when the angle between the particle's velocity vector **v** and the magnetic flux density vector **B** is an angle α. We resolve the velocity **v** into two components, i.e. a component parallel to **B** is

$$v_B = v \cos \alpha$$

and the component perpendicular to **B** is

$$v_{\perp B} = v \sin \alpha$$

The trajectory of the particle is then a 'helix', whose radius is

$$R_\alpha = \frac{mv_{\perp B}}{QB} = \frac{mv \sin \alpha}{QB}$$

and has the pitch

$$k = \frac{2\pi R_a}{v_{\perp B}} v_B = \frac{2\pi m v \cos \alpha}{QB} \tag{11.35}$$

(Refer to Figure 11.11.)

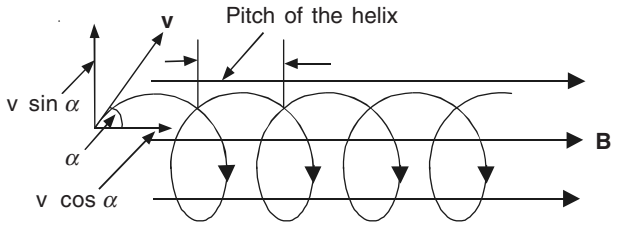


Figure 11.11 Helical path of a charged particle in a uniform **B** field.

The axis of the helix coincides with the direction of **B**. An example of this behaviour in (though non-uniform) **B** is that of the charged particles entering the earth's magnetic field from the outer space. These particles drift towards the north and the south poles, and the interaction between these particles and the air molecules in the upper layer of the atmosphere is believed to be the cause of the aurora.

11.5.4.3 The motion of a charged particle under the simultaneous action of the magnetic and the electric fields

When a charged particle is moving in a region where both the magnetic and the electric fields exist, the total force on the particle is the complete Lorentz force, and its equation of motion is

$$m\frac{d\mathbf{v}}{dt} = Q\mathbf{E} + Q(\mathbf{v} \times \mathbf{B}) \tag{11.36}$$

(a) *Parallel and uniform* **E** *and* **B**. We first consider the situation where both **E** and **B** are parallel (Figure 11.12). A charge Q is projected with a velocity **v** as shown in the figure, and experiences the force shown. The force due to the **E** field is (Q**E**) acting in the y-direction, i.e. $\mathbf{F}_y = Q\mathbf{E}$, which will cause the particle to be deflected in the y-direction (i.e. along **E**), by an amount y, i.e.

$$y = \frac{QEl^2}{2m\mathrm{v}^2} \tag{11.37a}$$

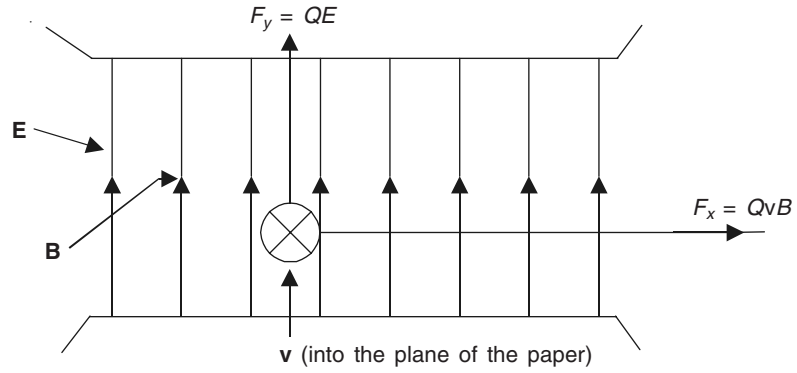


Figure 11.12 Charges in parallel **E** and **B**.

where l is the distance traversed in the z-direction.

The magnetic field **B** (also in the y-direction), produces the force $\mathbf{F}_x$ in the x-direction which is $F_x = Q\mathrm{v}B$, and this force causes a deflection in the x-direction which is given by

$$x = \frac{QBl^2}{2m\mathrm{v}} \tag{11.37b}$$

This is because the magnetic force only alters the direction of motion along a circular trajectory and not the speed, the radius of curvature of the path being given by Eq. (11.34). Eliminating v from the two above equations, we get

$$y = \left(\frac{2mE}{QB^2l^2}\right)x^2 \tag{11.38}$$

i.e. in a plane of given l (= the distance traversed in the z-direction), the particles of any velocity but with a definite (Q/m) will be distributed along a parabola.

(b) **E** *and* **B** *at right angles*. We consider a parallel plate diode with the static **B** parallel to the plates (Figure 11.13), and study the trajectory of an electron of charge ($-e$) and mass m, starting from the negative plate with the initial zero velocity, and from the point which we take as the origin of our coordinate system as shown in the figure.

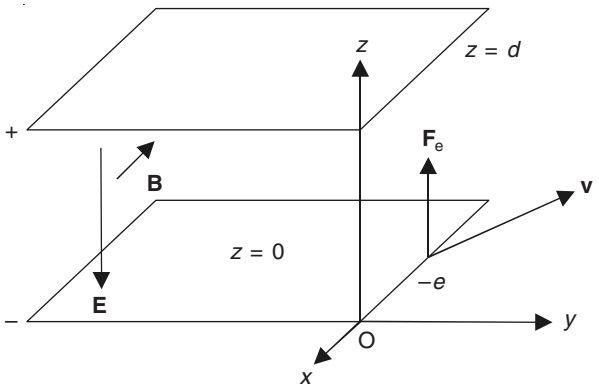


Figure 11.13 A parallel-plate diode with transverse magnetic field.

The equations of motion for the particle are

$$m\frac{dv_y}{dt} = ev_z B \tag{11.39}$$

and

$$m\frac{dv_z}{dt} = eE - ev_y B \tag{11.40}$$

Since $v_z = dz/dt$, integrating Eq. (11.39) with respect to time, we get

$$mv_y = ezB + C, \text{ where } C = 0, \text{ since } v_y = 0 \text{ at } z = 0$$

$\therefore$ $v_y = ezB/m$, and substituting in Eq. (11.40), we get

$$\frac{d^2z}{dt^2} + \left(\frac{e^2B^2}{m^2}\right)z = \frac{eE}{m} \tag{11.41}$$

Its solution is

$$z = C\sin\omega t + C\cos\omega t + \frac{eE}{m\omega^2}, \quad \text{where } \omega = \frac{eB}{m}$$

Using the initial condition that for $t = 0$, $z = 0$ and $dz/dt = 0$, we get

$$z = \left(\frac{eE}{m}\right)\frac{1 - \cos\left(\frac{eB}{m}\right)t}{\frac{e^2B^2}{m^2}}$$

and

$$y = \left(\frac{eE}{m}\right)\left[\frac{\left(\frac{eb}{m}\right)t - \sin\left(\frac{eB}{m}\right)t}{\frac{e^2B^2}{m^2}}\right] \tag{11.42}$$

These are the parametric equations of a cycloid—a combination of the linear motion and the circular motion.

The largest distance of the electron from the cathode is

$$z_{\max} = \frac{2\left(\frac{eE}{m}\right)}{\left(\frac{eB}{m}\right)^2} \tag{11.43}$$

and, hence the current through the diode exists only if

$$z_{\max} > d$$

i.e. if for a given E, the flux density **B** is such that

$$B > \sqrt{\frac{2Em}{ed}}, \qquad E = \frac{V_a}{d} \tag{11.44}$$

(V_a being the applied potential difference between the plates of the diode), then there will be no current through the diode. If B is smaller than the above quantity, all the electrons emitted would reach the anode. This is known as the 'magnetron effect'.

11.6 ENERGY STORAGE IN THE FIELD OF A COIL

We consider a coil of resistance R and self-inductance L (which is assumed to be constant, i.e. an air-cored coil). An emf $\mathcal{E}$ is applied to the terminals which gives a current I in the coil. The relationship for the coil is given by the equation

$$\mathcal{E} = RI + L\frac{dI}{dt}$$

We can rewrite this equation as

$$\mathcal{E}I = RI^2 + LI\frac{dI}{dt} = RI^2 + \frac{d}{dt}\left[\left(\frac{1}{2}\right)LI^2\right] \tag{11.45}$$

In the above equation, the left-hand-side term $\mathcal{E}I$ expresses the rate of energy input to the circuit, and RI^2 gives the rate of dissipation of the energy in the circuit resistance. This equation shows that a part of the energy input to the circuit gets dissipated in its resistance, and the rest increases the quantity $(1/2)LI^2$ which is called the '*stored energy*' of the coil. A practical demonstration of this stored energy is seen when a circuit is suddenly broken by opening the switch, an arc flashes across the switch contacts. The stored energy increases when I increases, and decreases when I decreases. In an alternating current circuit this energy is stored and taken out twice in each alternation.

This energy is being regarded as stored in the magnetic field, and that it is possible to regard a quantity of the stored energy to be associated with each unit volume of the field.

11.7 ENERGY STORAGE IN THE FIELD OF SEVERAL COILS

We consider a group of N coils with currents i_1, i_2, i_3, ..., i_N, and let their flux-linkages be Φ_1, Φ_2, Φ_3, Φ_4, ..., Φ_N, respectively. We wish to calculate the stored energy of the system, and so we suppose that the currents are built up gradually, such that at each instant each current has the same fraction k (< 1) of its final value. So when the currents are ki_1, ki_2, ki_3, ..., ki_N, respectively, the corresponding fluxes will be $k\Phi_1$, $k\Phi_2$, $k\Phi_3$, $k\Phi_4$, ..., $k\Phi_N$. In a time interval δt, the currents and the fluxes are increased by an additional fraction δk.

Then the induced emf in the circuit 1, say, is $\Phi_1 \delta k/\delta t$, which is opposite in direction to the current ki_1.

To maintain this current, the energy source in the circuit 1, must supply to it (the circuit1), at the rate which is given by

$$(ki_1)\,\Phi_1 \frac{\delta k}{\delta t}$$

$\therefore$ The energy supplied during this time-interval δt is $= i_1\Phi_1 k\,\delta k$.

$\therefore$ The total energy supplied to all the N coils during this time-interval δt is

$$k\delta k\,(i_1\Phi_1 + i_2\Phi_2 + i_3\Phi_3 + \cdots + i_N\Phi_N)$$

$\therefore$ The total energy supplied during the whole process of the current build-up is

$$\left\{\int_0^1 (k\,\delta k)\right\} \Sigma\, i\Phi = W$$

$$\therefore \qquad W = \left(\frac{1}{2}\right)\Sigma\, i\Phi \tag{11.46}$$

This expression has been derived for the current-carrying wires which are infinitely thin, so that Φ is exactly defined. With wires of finite thickness, the current in each wire can be divided into thin filaments of current, and thus calculate the stored energy using each filament as a separate circuit.

Now, we express the energy W in terms of the self- and the mutual-inductances of the circuits.

We note that Φ_1 due to i_1 only $= L_{11}i_1$,

and that Φ_1 due to i_2 only $= M_{12}i_2$.

$\therefore$ The flux due to all the currents in all the N circuits

$$\begin{aligned}
\Phi_1 &= L_{11}i_1 + M_{12}i_2 + M_{13}i_3 + \cdots + M_{1N}i_N \\
\Phi_2 &= M_{21}i_1 + L_{22}i_2 + M_{23}i_3 + \cdots + M_{2N}i_N \\
&\;\;\vdots \qquad\qquad \vdots \quad \vdots \quad \vdots \\
\Phi_N &= M_{N1}i_1 + M_{N2}i_2 + M_{N3}i_3 + \cdots + L_{NN}i_N
\end{aligned} \tag{11.47}$$

Note that $M_{nm} = M_{mn}$ (a general proof of this reciprocal property based on the stored energy follows immediately).

Multiplying each flux by the corresponding current, i.e. Φ_1 by i_1, Φ_2 by i_2, Φ_3 by i_3, ⋯, Φ_N by i_N, and adding up, the expression for the stored energy is

$$W = \left(\frac{1}{2}\right)\left[L_{11}i_1^2 + L_{22}i_2^2 + \cdots + L_{NN}i_N^2 + 2M_{12}i_1i_2 + 2M_{13}i_1i_3 + \cdots\right] \tag{11.48}$$

The expression for a single coil is a particular case of this general expression.

11.7.1 Reciprocal Property of Mutual Inductance

This property for the two circuits was stated in the Section 10.6 without giving a proof for it . So, now we prove it for the two circuits with mutual inductance, by considering the energy in them. Let the currents i_1 and i_2 (in the two circuits) be established in them, first by bringing the current in the circuit (1) to the value i_1, and then the current in the circuit (2) to the value i_2.

The establishment of the current i_1 requires the energy, i.e.

$$W = \left(\frac{1}{2}\right)L_{11}i_1^2$$

from the source in the circuit (1).

During the establishment of i_2, i_1 remains constant, and so, there is no change of the flux in the circuit (2).

∴ The energy required from the source in the circuit (2) is simply the amount associated with i_2, i.e.

$$W_2 = \left(\frac{1}{2}\right)L_{22}i_2^2$$

While the current in the circuit (2) (let us call it i_2' during the intermediate stage) is increasing to its final value i_2, it induces in the circuit (1), an emf of value $M_{12}(di_2'/dt)$. With the positive mutual inductance, this emf makes it more difficult for the current i_1 in the circuit (1) to be maintained at that value. Its maintenance requires an extra amount of energy from the source in the circuit (1), this being given by

$$W_3 = \Sigma\, i_1 M_{12}\left(\frac{di_2'}{dt}\right)dt = M_{12}i_1i_2$$

∴ The total energy in the system, when both the currents have reached their final values, is

$$W = W_1 + W_2 + W_3 = \left(\frac{1}{2}\right)L_{11}i_1^2 + \left(\frac{1}{2}\right)L_{22}i_2^2 + M_{12}i_1i_2 \tag{11.49}$$

If the currents i_1 and i_2 had been built in the reverse order, then

$$W' = \left(\frac{1}{2}\right)L_{22}i_2^2 + \left(\frac{1}{2}\right)L_{11}i_1^2 + M_{21}i_2i_1 \tag{11.49′}$$

Now, $W = W'$, otherwise by storing the energy in one order, and depleting it in another order, we would have a perpetual source of energy.

∴ $M_{12} = M_{21}$ (This is same as in Section 10.6.)

This is true for the air-cored circuits, and may not be generally true for the iron-cored circuits.

11.8 DISTINCTION BETWEEN THE POTENTIAL ENERGY AND THE STORED ENERGY

The expression for the potential energy is

$$U = - I\Phi \tag{11.10}$$

and the expression for the stored energy is

$$W = \left(\frac{1}{2}\right)\Sigma\, I\Phi \tag{11.46}$$

As an example, we consider a pair of circuits carrying the currents I_1 and I_2, respectively, in the same direction (Figure 11.14).

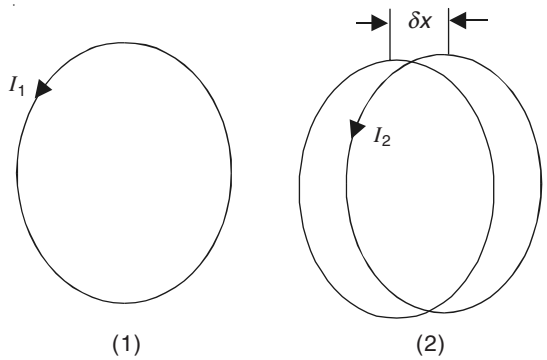


Figure 11.14 Energy in two circuits.

The potential energy in the field of either circuit is given by

$$U = - M_{12}I_1I_2 \tag{11.49a}$$

and the stored energy in the field is given by

$$W = \left(\frac{1}{2}\right)L_{11}I_1^2 + \left(\frac{1}{2}\right)L_{22}I_2^2 + M_{12}I_1I_2 \tag{11.49b}$$

Let the circuit (2) be moved by a distance δx, so that the mutual inductance between the two circuits increases by δM_{12}. The potential energy of the circuit then decreases by $\delta M_{12}I_1I_2$ and this energy is transformed into the mechanical work done by the force acting on the circuit (2). At the same time, the potential energy of the circuit (1) in the field of the circuit (2) is also reduced by $\delta M_{12}I_1I_2$. And this is balanced by the energy stored in the field.

The storing places of the potential energy are the current sources in the circuits. An equal quantity is withdrawn from each source, but only half the amount withdrawn is used in doing the mechanical work, the rest being stored in the field.

In fact, there are two sources of the energy: the current source, and the field.

Everything is so arranged that for every unit of energy (joule) taken from the source of the potential energy for the external consumption, a second unit from this source is transferred to the storage in the field. On the other hand, if the extraneous forces pull the circuits apart, then the current sources are enabled (through the withdrawal of the energy from the field), to receive energy equal to twice the work done by the forces. In fact, the 'stored energy' is only another

form of the potential energy. If we are to speak more accurately and rigorously, what we should say is:

"The potential energy of a group of current-carrying circuits takes two forms: the current-source-stored energy U, and the field-stored energy W".

11.9 FORCES BETWEEN THE TWO CIRCUITS IN TERMS OF THE MUTUAL INDUCTANCE

From Eq. (11.49b), the potential energy of a circuit (2) in the field of a circuit (1) is given by

$$U = - M_{12} I_1 I_2$$

$\therefore$ From Eq. (11.17),

$$F_x = - \frac{\partial U}{\partial x} = I_1 I_2 \left(\frac{\partial M_{12}}{\partial x} \right) \tag{11.49c}$$

Hence, if the mutual inductance between the two circuits is known, then the force between them for various spacings can be found out.

11.10 STORED ENERGY IN TERMS OF THE MAGNETIC FIELD VECTORS

We start from the expression for the stored energy, which is

$$W = \left(\frac{1}{2} \right) \Sigma \, i\Phi$$

We now consider a closed filament bounded by the lines of force, containing a flux $\delta\Phi$. The contribution of this flux to the stored energy is given by

$$\delta W = \left(\frac{1}{2} \right) \delta\Phi \, \Sigma \, i \tag{11.50}$$

where $\Sigma \, i$ = all the currents linking with $\delta\Phi$ (Figure 11.15).

$$= \oint_C \mathbf{H} \cdot d\mathbf{l} \qquad \text{(by Ampere's law)}$$

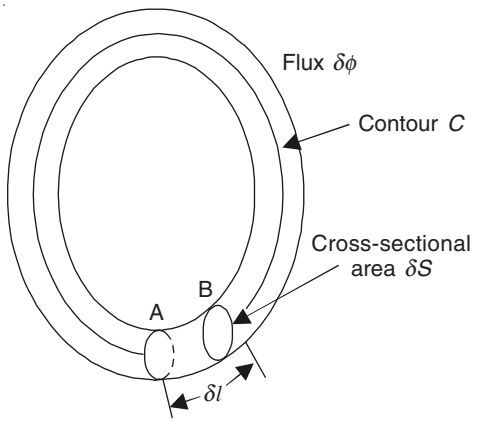


Figure 11.15 A closed filament.

$$\therefore \qquad \delta W = \left(\frac{1}{2}\right)\delta\Phi \oint_C \mathbf{H}\cdot d\mathbf{l}$$

$$= \left(\frac{1}{2}\right)\delta\Phi\,\Sigma\,\mathbf{H}\cdot d\mathbf{l}$$

Also, $\delta\Phi = B\delta S$.

$$\therefore \qquad \delta W = \left(\frac{1}{2}\right)\Sigma\,[BH\,(\delta S\delta l)] = \left(\frac{1}{2}\right)\Sigma\,(BH\delta v) \tag{11.51}$$

where δv is the volume element = $\delta S\delta l$.

$\therefore$ Over the whole volume, $W = \left(\frac{1}{2}\right)\iiint (BH\,dv)$ (11.52)

Thus the energy appears to be distributed all over the field, the energy density at any point being [(1/2)(*BH*)] joules/metre^3. We compare this with the similar relation for the electric field, i.e.

$$W_e = \left(\frac{1}{2}\right)(DE\,dv)$$

In this derivation, it was assumed that the flux increased linearly with the current, which is true for all media with constant permeability. But this is not true for the ferromagnetic materials, like iron.

11.11 ENERGY STORAGE IN A REGION CONTAINING IRON

From the previous section, we now generalize the expression for the energy storage for iron, where μ is not constant. We consider an iron ring of mean circumference l, and cross-section S. Initially it is in a demagnetized state. It is magnetized by a current I through a uniform toroidal winding of N turns. At an intermediate stage of magnetization, in a time-interval δt, the current increases by δI, and the flux by $\delta\Phi$.

$\therefore$ The induced emf in this interval = $N\left(\frac{\delta\Phi}{\delta t}\right)$

$\therefore$ To maintain the current I, the extra energy needed is

$$\delta W = IN\left(\frac{\delta\Phi}{\delta t}\right)\delta t = NI\,\delta\Phi$$

$\therefore$ Over the whole process, the total energy required is

$$W = \int_0^{\Phi} NI\,\delta\Phi$$

The mean values of B and H in the iron are given by $B = \Phi/S$ and $H = NI/l$. Hence writing the last equation in terms of B and H, we get

$$W = lS\int_0^B H\,dB = \int_0^B H\,dB \times \text{volume of iron} \tag{11.53}$$

Though this relation has been proved for an iron ring, it is true, in general.

$$\therefore \qquad \text{Energy density} = \int_0^{B_S} H dB \text{ joules/m}^3 \qquad (11.54)$$

The significance of these equations can be better appreciated with reference to an actual *B–H* curve (Figure 11.16). The iron starts from the demagnetized condition, which is represented by the point O and is brought up to the flux density B_S at the point S. Equation (11.54) states that the energy storage density is equal to the shaded area ($OKNSB_SO$). If however the permeability of the iron was assumed to be constant, then the energy density would be given by the triangle (OSB_S). If in a magnetic circuit, a part is air and a part is iron, then the value of *H* in iron is much less than in air. Hence the density of the magnetic energy is much greater in air than in iron.

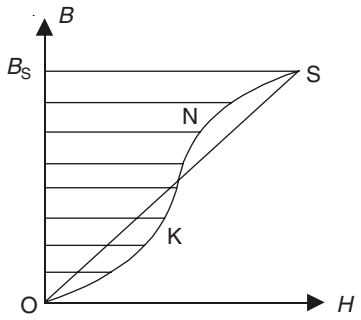


Figure 11.16 *B–H* curve for iron.

11.12 ATTRACTIVE FORCE BETWEEN MAGNETIZED IRON SURFACES

We consider a U-shaped electromagnet attracting an iron plate as shown in Figure 11.17. Let the gap between the surfaces be increased by δx. If, during this change, **B** is maintained at its constant value, then the stored energy in the iron does not change, but that in the air-gap increases by $(1/2)BHA\,\delta x$. This extra energy is supplied by the work done by the mechanical force **F** which caused the gap to increase.

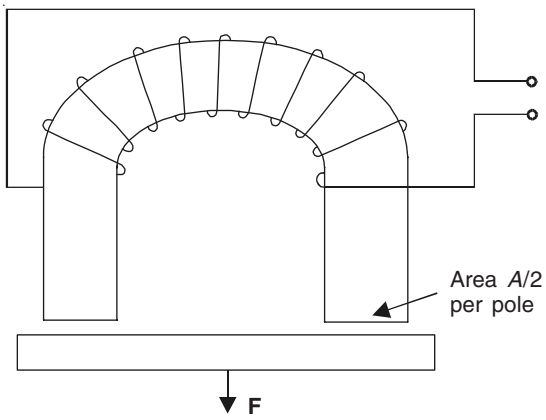


Figure 11.17 U-shaped electromagnet.

$$\therefore \qquad F\delta x = \left(\frac{1}{2}\right) BHA\ \delta x \rightarrow \frac{F}{A} = \left(\frac{1}{2}\right) BH = \frac{B^2}{2\mu_0}$$

$$\therefore \qquad \text{Force per unit area} = \frac{B^2}{2\mu_0} \text{ newtons/metre}^2 \tag{11.55}$$

Note that, if B is 1 W/m^2, the force will be 4 kg/cm^2. It must be noted that the force on the iron depends on the *local* value of B, and not on the distant apparatus by which B is produced. In practice, the current in the electromagnet remains constant as the plate is pulled away, because as the air-gap increases, the flux density both in the iron and the air-gap, decreases. In such problems, the assumption of the constant flux is the simplest approach, as then there is no induced emf in any winding, and hence no exchange of energy between the current sources and the field.

11.12.1 Forces on Circuits with Associated Iron

We consider, now, a circuit C which carries a current i, and is situated in a magnetic field containing other circuits and iron. When there is a force F_x on the circuit, in the direction x, causing the circuit to move a distance δx, the flux-linkage increases by an amount $\delta\Phi$, and a consequent increase in the stored energy of the field by an amount δW. So the current source has to supply an amount of extra energy $i\delta\Phi$, a part of which is stored in the field and a part of it being converted to the mechanical work, i.e.

$$i\delta\Phi = F_x\,\delta x + \delta W$$

$$\therefore \qquad F_x = i\frac{\partial\Phi}{\partial x} - \frac{\partial W}{\partial x} \tag{11.56}$$

Similarly in a rotation $\delta\theta$, the torque is given by

$$T_\theta = i\frac{\partial\Phi}{\partial\theta} - \frac{\partial W}{\partial\theta} \tag{11.57}$$

These equations are useful for solving the problems related to the electrical machines, where the circuits are of composite types as mentioned here.

11.13 ENERGY LOSS ASSOCIATED WITH HYSTERESIS IN IRON

In Section 11.11, we have seen that the energy stored in a volume δv of the magnetized iron is

$$\delta W = \delta v \int_0^{B_1} H\ dB \tag{11.53}$$

In this expression, the energy input is calculated as B is increased from zero to a value B_1, the implicit assumption being that when the magnet was demagnetized, the initial condition was that $B = 0$ and $H = 0$. But this condition is not always correct due to the hysteresis characteristic of the material. So, we generalize the above expression for a change of the flux density level from B_1 to B_2 in the material, the required energy for the change being

$$\delta W_{12} = \delta v \int_{B_1}^{B_2} H\ dB \tag{11.58}$$

Now, this formula is applied to a specimen which is undergoing cyclic changes of magnetization according to the hysteresis cycle shown in Figure 11.18. If B changes from the negative value at the point P to the positive value at R, then the energy input necessary for this change is

$$\delta W_{PR} = \delta v \int_{P}^{R} H\, dB = \delta v \times \text{Area PQRS}$$

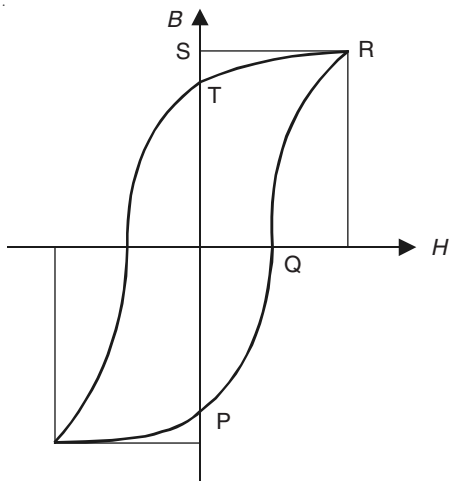


Figure 11.18 Hysteresis loop of iron.

In the next step, when B is reduced from the point R to the point T, the energy recovered from the specimen is

$$\delta W_{RT} = \delta v \int_{R}^{T} H\, dB = -\,\delta v \times \text{Area RST}$$

since during the traverse from R to T, dB is negative. Adding these two steps, the reversal of B from its negative value at P to its positive value at T has been accompanied by an energy input of

$$\delta W_{PR} + \delta W_{RT} = \delta v \times (\text{Area PQRS} - \text{Area RST})$$

$$= \delta v \times \text{Area PQRT} \qquad (11.59)$$

It should be noted that if the material had no hysteresis, then the energy input for such a reversal of B (from $-B_1$ to $+B_1$) would have been zero. This energy required by the hysteretic material is not recoverable. If B is changed from its positive value at T to the equal negative value at P, it will be seen that an equal amount of energy as given by Eq. (11.59) is drawn in. So we find that

$$\text{Energy input per cycle of magnetization} = \delta v \times \text{Area of the hysteresis loop} \qquad (11.60)$$

The area must be calculated in terms of the SI units of B and H, and the result gives the energy input in joules per metre^3 per cycle. This energy is converted into heat and represents a loss which is known as the 'hysteresis loss'. This energy cannot be called the stored energy as it is not recoverable. This loss is the energy transformed into heat in periodically orienting the magnetized domains between which a kind of friction exists. C.P. Steinmetz observed that for a variety of

iron, and for a range of flux density, the area of the loop was empirically proportional to $(B_{max})^n$, so that

$$\text{Loss per metre}^3 \text{ per cycle} = \lambda_h (B_{max})^n \text{ joules} \qquad (11.61)$$

where λ_h and n are the material characteristics. However $n \simeq 1.6$ for iron, mild steel, silicon steel, and cast iron. The other constant λ_h which is known as the Steinmetz constant, varies from 500 for silicon steel to 3000 for the cast iron.

The law which we have discussed so far, applies to the situations where B purely alternates in magnitude, but does not change its direction. In electrical machines and transformer corners, the direction of B rotates as well as the magnitude changes. Such a process is associated with a power loss known as the 'rotational hysteresis loss', and this loss does not follow the Steinmetz law.

11.14 MAXWELL'S STRESSES IN THE MAGNETIC FIELD

Let us consider the force between two bar magnets A and B, as shown in Figure 11.19. We shall now see, if we can calculate the forces between them, by considering a system of stresses on any closed surface round either of the magnets. Such an approach has been found to be very convenient in a number of practical problems, such as the design of the structures of electrical machines.

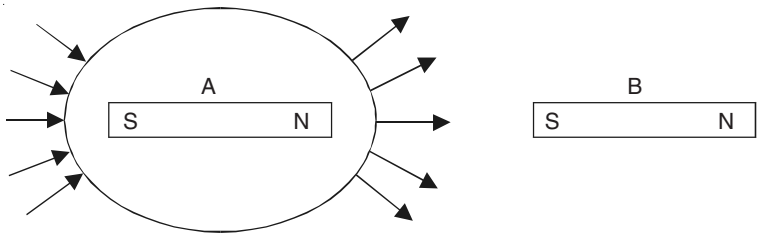


Figure 11.19 Force between two bar-magnets and stresses in the field space.

We have already seen in Section 11.12, that when the magnetic flux enters a piece of iron normally, there is an outward normal force on the iron surface at the rate of $B^2/2\mu_0$ where B is the local value of the flux density. So we can argue that any stress in the direction of the lines of force must be the tensile stress of magnitude $B^2/2\mu_0$. But this is not the complete picture. When we have two parallel wires carrying currents in the opposite directions, there is a repulsive force between them which is at right angles to the flux lines (which are circles surrounding the wires). This is equivalent to a compressive stress at right angles to the lines of force. To find the magnitude of this stress, let us consider two concentric hollow tubes of radii a and b respectively ($a < b$) and the currents $\pm I$ as shown in Figure 11.20.

$\therefore$ $\quad B = \dfrac{\mu_0 I}{2\pi r}$ $\quad$ for $a < r < b$

$\quad B = 0$ $\quad$ for $r < a$ and $r > b$

$\therefore$ Just inside the outer tube, $\quad B = \dfrac{\mu_0 I}{2\pi b}$

and just outside it, $\quad B = 0$

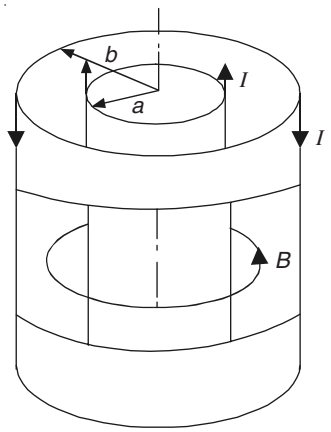


Figure 11.20 A portion of concentric tubes carrying current $+I$ and $-I$ in opposite directions.

$\therefore$ The mean value of B inside the wall of the outer tube $= \dfrac{\mu_0 I}{4\pi b}$

The current per unit length of the circumference $= \dfrac{I}{2\pi b}$

$\therefore$ The pressure on the outer tube $= \dfrac{\mu_0 I}{4\pi b} \times \dfrac{I}{2\pi b}$

$$= \frac{B^2}{2\mu_0} \tag{11.62}$$

Hence we can say that the force between the magnetic bodies separated by an air-space can be considered to be transmitted by a system of stresses, which are:

1. A tensile stress $B^2/2\mu_0$, in N/m^2, along the line of force at each point (normal stress)
2. A compressive stress $B^2/2\mu_0$, in N/m^2, in every direction at right angles to the line of force at each point (shear stress).

(11.63)

These stresses are valid not only for the air-space but also for all non-magnetic media. For the ferromagnetic media, Maxwell derived the following system of stresses (the only assumption being negligible hysteresis):

1. A tensile stress $[BH - \mu_0 H^2/2]$, in N/m^2, along the line of force at each point (normal stress)
2. A compressive stress $[\mu_0 H^2/2]$, in N/m^2, in every direction at right angles to the line of force at each point (shear stress).

(11.64)

These are known as the Maxwell's stresses, and are completely general except for the hysteresis assumption.

Equations (11.64) can also be rewritten as:

1. A tensile stress BH, in N/m^2, along the line of force at each point
2. A compressive stress $\mu_0 H^2/2$, in N/m^2, in every direction at each point.

11.15 A DEFINITION OF THE SELF-INDUCTANCE IN TERMS OF THE STORED ENERGY

Inductances, both self- and mutual-, have been defined in Chapter 10, Section 10.6, in terms of the flux-linkage per unit current. But there were difficulties in calculating the self-inductance of some configurations, such as the parallel wire circuit. So, now, we try to find a new definition of the self-inductance in terms of the stored energy. Let us note that the stored energy associated with a coil of inductance L and carrying a current I is

$$W = \left(\frac{1}{2}\right) LI^2 \qquad \text{[Eq. (11.45)]}$$

Also, the expression for the stored energy W in terms of the flux and the current, applicable to the wires of finite dimensions, was obtained as

$$W = \left(\frac{1}{2}\right) (\Sigma\, \Phi \delta i)$$

$$\therefore \qquad L(\text{self-inductance}) = \left(\frac{1}{I^2}\right) \Sigma\, \Phi \delta i \tag{11.65}$$

where I stands for the total current = $(\Sigma\, \delta i)$.

In this case, both the current and the flux can be divided into closed tubes such that each tube contains the same flux $\delta\Phi$, and each current tube the same current δi. Then each linkage of a flux tube and a current tube contributes $(\delta\Phi\, \delta i)$ to the quantity $(\Sigma\, \Phi \delta i)$. It is possible that both the tubes may form complicated spirals.

But, in fact, one may concentrate on a particular δi tube, and count the linkages of all the flux tubes to obtain $\Phi \delta i$, or vice versa, so that the alternative expression will be:

$$L\ (\text{self-inductance}) = \left(\frac{1}{I^2}\right) \Sigma\, i \delta\Phi \tag{11.66}$$

i.e. divide the flux into the elements, and count the current tubes linking it. It should be noted that there may be cases where i can exceed the total current I, depending on the convolutions round a flux tube.

Furthermore, since the stored energy can be expressed in terms of the field vectors, the self-inductance can be expressed as

$$L = \left(\frac{1}{I^2}\right) \iiint_v \mathbf{B} \cdot \mathbf{H}\, dv \tag{11.66a}$$

and the mutual inductance between the two circuits as

$$M_{12} = \left(\frac{1}{I_1 I_2}\right) \iiint_v \mathbf{B}_1 \cdot \mathbf{H}_2\, dv \tag{11.66b}$$

11.15.1 Internal Self-inductance of a Straight Cylindrical Conductor

When the self-inductances of the circuits were calculated, say, in Section 10.6.2, the contribution to L was from the flux which did not cross the cross-section of the current-carrying conductors. So we call such Ls as the 'external self-inductance'. When the contributions from the flux traversing

the conductor cross-section are considered, the inductance is called the 'internal self-inductance'. The inductance of the parallel wire circuit, obtained in Section 10.6.2, as

$$\left(\frac{\mu_0}{\pi}\right)\ln\left(\frac{b}{a}\right) \qquad \text{H/m}$$

was the external self-inductance of the circuit.

Now, we consider a straight circular cylindrical conductor of radius a, carrying a current I which is uniformly distributed over its cross-section. The parallel return conductor for the current is so far away, i.e. $b >> a$, that the flux lines within the conductor cross-section can be considered as concentric circles (Figure 11.21).

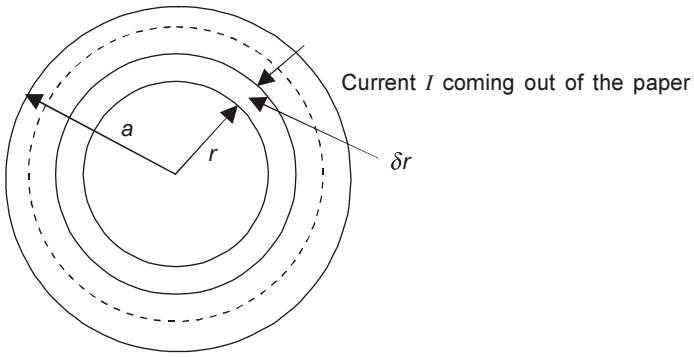


Figure 11.21 A circular conductor.

The flux density within the conductor at a radius r, is

$$B = \frac{\mu_0 I}{2\pi a^2} r$$

We use the formula (11.65), and divide the internal flux into the rings of radius r and width δr.

$$\therefore \qquad \delta\Phi = B\delta r = \frac{\mu_0 I r \delta r}{2\pi a^2} \text{ per unit axial length}$$

The current linked by this flux filament

= the current within the cylinder of radius r

$$= \frac{Ir^2}{a^2}$$

$$\therefore \qquad i\delta\Phi = \frac{\mu_0 I^2}{2\pi a^4} r^3 \delta r$$

$$\therefore \quad \text{The internal self-inductance, } L_i = \left(\frac{1}{I^2}\right)\Sigma\, i\delta\Phi$$

$$= \frac{\mu_0}{2\pi a^4}\int_0^a r^3 dr$$

$$= \frac{\mu_0}{8\pi} \text{ H/m of the conductor} \qquad (11.67)$$

and hence for the parallel circuit, $L_{iC} = \dfrac{\mu_0}{4\pi}$

and we have $$L_e = \left(\frac{\mu_0}{\pi}\right)\ln\left(\frac{b}{a}\right)$$

$\therefore$ The total self-inductance, $L = L_e + L_{iC}$

$$= \left(\frac{\mu_0}{\pi}\right)\left[\ln\left(\frac{b}{a}\right) + \frac{1}{4}\right] \tag{11.68}$$

It should be noted that the internal self-inductance of the conductor is independent of the radius of the conductor.

11.16 ENERGY AND FORCES IN ELECTROMECHANICAL SYSTEMS

An electromechanical device (i.e. an electrical motor or generator or any other transducer, in general) is a link between an electrical and a mechanical system. The coupling between the two systems is through the medium of the fields of the electric charges. Both the electric and the magnetic fields are present in general, and the energy storage in these fields is associated with the energy conversion. During the process of energy conversion, the energy in the coupling field may change or tend to change. The electromechanical energy conversion thus depends on the existence of the phenomena interrelating the magnetic and the electric fields on one hand and the mechanical force and the motion on the other.

We have found from our previous discussion in Section 11.8 on 'Distinction between Potential energy and Stored energy', that when the potential energy in a system is used for any mechanical work, out of the total energy taken, only half is used for the mechanical work and the other half is stored in the field. So remembering that the potential energy of the system itself is in the electrical sources, we can now write down the energy balance equation. However we restrict our treatment to the quasi-static (or low frequency) systems and the fields which give rise to the forces in a particular element are either electric or magnetic, but not both. So we can consider separately the forces due to the electric fields and the forces due to the magnetic fields. In the systems under consideration, there are four important forces of electrical origin, i.e.

1. the force due to an electric field acting on a free charge;
2. the force due to an electric field acting on a polarizable material;
3. the force due to a magnetic field acting on a moving free charge (current);
4. the force due to a magnetic field acting on a magnetizable material.

So we shall now write down the energy balance equation, using the principle of conservation of energy, for a magnetic field electromechanical system (Figure 11.22).

Increase in the potential energy (= Energy input from the electrical source)

= Mechanical energy output (+ friction and windage losses) + Increase in the energy stored in the coupling magnetic field (+ associated losses)

or

$$dW_{elec} = dW_{mech} + dW_{fld} \tag{11.69}$$

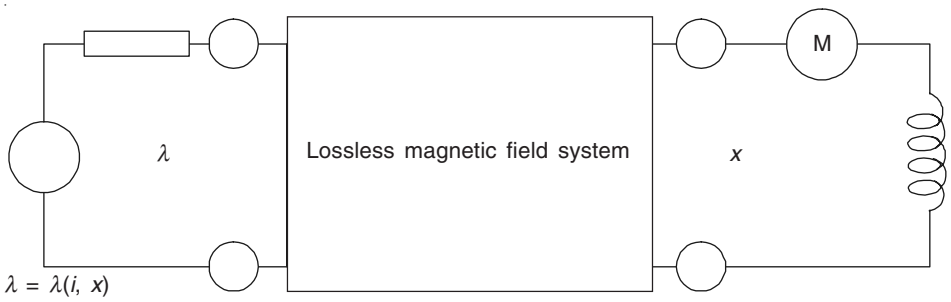


Figure 11.22 A terminal characteristic representation of a magnetic field electromechanical system.

Now, the change in the potential energy, or the electrical energy input

$$= dW_{\text{elec}} = e\, i\, dt \tag{11.70}$$

where

e = the time-rate of change of flux linked by the electrical circuit in the system
i = the current in the system.

If we use the notation λ—to denote the flux-linkage (so as to distinguish from Φ—which is used as the notation for the magnetic flux in a region), then

$$e\, dt = d\lambda$$

$$\therefore \qquad dW_{\text{elec}} = i\, d\lambda \tag{11.71}$$

In general, the flux linked by the electrical circuit is a function of the current i and the displacement x, i.e.

$$\lambda = \lambda\,(i, x) \tag{11.72}$$

Also, when we know i and x of the magnetic field system, the value of the flux-linkage λ would be known.

We next make an assumption about the mechanical force of the electrical origin and say that for a given current i and position x, the force is single-valued and has the form

$$F_{\text{e}} = F_{\text{e}}\,(i, x) \tag{11.73}$$

This is justifiable if these variables are the ones controlling the stored energy in the field.

We define F_{e}, the mechanical force of electrical origin applied to the mechanical node of the system, in a direction so as to increase the displacement x.

$$\therefore \qquad dW_{\text{mech}} = F_{\text{e}}\delta x \tag{11.74}$$

From Eq. (11.72), we can solve for the current i in terms of the flux-linkage λ and the displacement x, i.e.

$$i = i(\lambda, x) \tag{11.75}$$

Hence F_{e} can also be expressed as a function of the magnetic field in which the energy is stored, i.e.

$$F_{\text{e}} = F_{\text{e}}\,(\lambda, x) \tag{11.76}$$

However, it should be noted that F_{e} is not the same function in the two Eqs. (11.73) and (11.76).

Though in our present analysis, we are considering a system with one electrical and one mechanical terminal pair, this discussion can be generalized to any arbitrary number of terminal pairs.

We can now consider the energy of the system. For the simplicity of the analysis, we consider the coupling network to be lossless, i.e. the energy fed into the system by the electrical and the mechanical terminal pairs is stored in the magnetic field and is completely recoverable, i.e. it is a 'conservative system'. Since we are considering the magnetic field storage, we use the notation W_m, i.e. $W_{fld} = W_m$ and the conservation of the power of the system can be written as

$$\frac{dW_m}{dt} = i\frac{d\lambda}{dt} - F_e\frac{dx}{dt} \tag{11.77}$$

In this equation

$\frac{dW_m}{dt}$ = the time-rate of increase of the stored magnetic energy

$i\frac{d\lambda}{dt}$ = the electrical power input at the electrical terminals of the system

$-F_e\frac{dx}{dt}$ = the power input at the mechanical terminals.

The negative sign of the mechanical power is due to the definition of F_e which is defined to act on (i.e. into) the mechanical node. We multiply this equation by the time element δt to obtain the equation for the conservation of energy, i.e.

$$dW_m = i\,d\lambda - F_e dx \tag{11.78}$$

Next, we now wish to obtain the Force-Energy relations. From Eqs. (11.75), (11.76), and (11.78), it is clear that the magnetic energy W_m can be expressed as a function of the two independent variables λ and x; i.e.

$$W_m = W_m(\lambda, x) \tag{11.79}$$

Since we are considering a conservative system, the energy must be a single-valued function of the independent variables (λ, x) with finite second partial derivatives. So we can take the total differential of the field energy W_m and we get

$$dW_m = \left(\frac{\partial W_m}{\partial \lambda}\right)d\lambda + \left(\frac{\partial W_m}{\partial x}\right)dx \tag{11.80}$$

Subtracting this equation from Eq. (11.78), we get

$$0 = \left(i - \frac{\partial W_m}{\partial \lambda}\right)d\lambda - \left(F_e + \frac{\partial W_m}{\partial x}\right)dx \tag{11.81}$$

Since the variables x and λ are independent, the required condition to hold the above equation is given by

$$i = \frac{\partial}{\partial \lambda}\left[W_m(\lambda, x)\right]$$

$$F_e = -\frac{\partial}{\partial x}\left[W_m(\lambda, x)\right] \tag{11.82}$$

∴ If the stored energy is known, the electrical and the mechanical terminal relations can be calculated.

These equations can be generalized for a system with an arbitrary number of electrical and mechanical (translational and/or rotational) terminal pairs.

11.16.1 Force–(Co-energy) Relations of the Electromechanical System

In the magnetic field problem, the flux-linkage λ is used as the independent variable, while the current is described by the terminal relation. This is a natural choice because of the form of the conservation of energy equation [i.e. Eq. (11.78)]. However it should be possible to analyze such systems using the current i as an independent variable, which is a more convenient choice. We start with the same conservation of energy equation, i.e.

$$dW_m = i\,d\lambda - F_e dx \tag{11.78}$$

and since i is the independent variable now,

$$\lambda = \lambda(i, x) \tag{11.72}$$

$$F_e = F_e(i, x) \tag{11.73}$$

These are our terminal relations.

By the rules of differentiation

$$d(\lambda i) = id\lambda + \lambda di$$

and Eq. (11.78) becomes

$$dW_m = d(\lambda i) - \lambda di - F_e dx$$

$$\therefore \quad d(\lambda i) - dW_m = dW'_m = \lambda di + F_e dx \tag{11.83}$$

where

$$dW'_m = d(\lambda i) - dW_m \tag{11.84}$$

and is called the co-energy of the system.

Since

$$W'_m = W'_m(i, x) \tag{11.85}$$

$$\therefore \quad dW'_m = \left(\frac{\partial W'_m}{\partial i}\right) di + \left(\frac{\partial W'_m}{\partial x}\right) dx \tag{11.86}$$

Subtracting Eq. (11.86) from Eq. (11.83), we get

$$0 = \left(\lambda - \frac{\partial W'_m}{\partial i}\right) di + \left(F_e - \frac{\partial W'_m}{\partial x}\right) dx \tag{11.87}$$

Since di and dx are independent,

$$\lambda = \left(\frac{\partial}{\partial i}\right)\left[W'_m(i, x)\right]$$

and

$$F_e = \left(\frac{\partial}{\partial x}\right)\left[W'_m(i, x)\right] \tag{11.88}$$

∴ If the stored energy, and hence the co-energy are known, the electrical and the mechanical terminal relations can be calculated. Also, as before, these equations can be generalized for a system with any arbitrary number of electrical and mechanical terminal pairs.

Note: This manipulation of the conservation of energy in terms of the new independent variables, is called a 'Legendre Transformation' in Classical Mechanics and Thermodynamics.

The energy and the co-energy relations can be easily appreciated graphically by studying the flux–mmf characteristic. Figures 11.23(a) and (b) show this relationship for both the linear and the nonlinear cases.

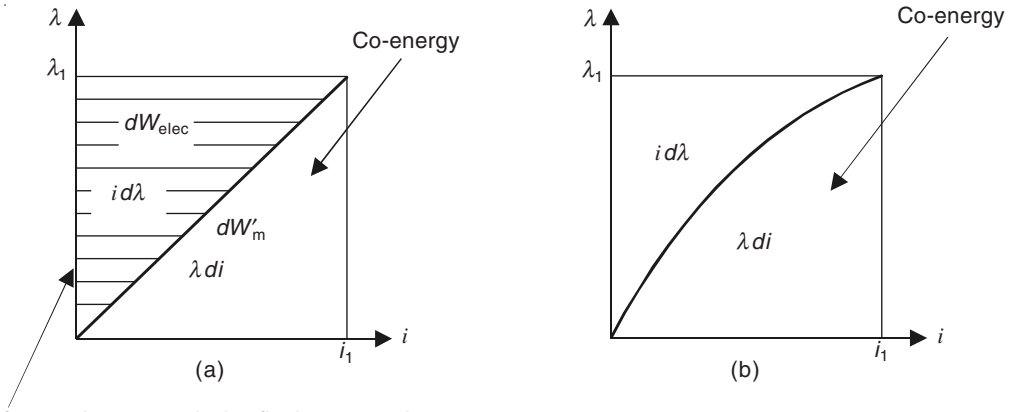


Figure 11.23 Flux–mmf characteristic and magnetic co-energy: (a) linear and (b) nonlinear.

Note: $dW_{elec} = id\lambda$

$= dW_{fld}(= dW_m) + F_e dx$

If the configuration is fixed, only then

$$dW_{elec} = dW_{fld} + 0 = dW_m$$

11.17 STABILITY OF MAGNETIC SYSTEMS

It has been shown in Section 3.8, that a static system consisting of particles, which either repel or attract one another with forces varying inversely with their distances, is basically unstable. This was proved by Earnshaw (as mentioned earlier). It had been applied by Maxwell to the electrostatic field, and by L. Tonks to the magnetostatic fields. The most complete theory was given by Braunbek in 1939. He proved that the levitation in electrostatic, magnetostatic, and stationary electromagnetic fields is impossible, because of instability, if the relative permittivity ε_r and the relative permeability μ_r are ≥ 1 everywhere in the system. Since $\varepsilon_r \geq 1$, for all materials under all circumstances, it follows that levitation is impossible in an electrostatic field, whereas it is at least not impossible if the field is magnetic and materials with $\mu_r < 1$ are present. There are two classes of such materials, i.e. diamagnetic materials whose μ_r is slightly less than unity, and materials in the superconductive state for which $\mu_r \simeq 0$. Furthermore it should also be appreciated that Earnshaw's theorem deals with potential energy *only*. For those systems, which possess other

types of energy as well, such as the gravitational system of sun and the planets possessing both kinetic and potential energy, stable equilibrium is possible. So also, by a suitable interaction of potential and kinetic energy, it is possible to suspend objects in stable equilibrium (i.e. levitate) in electromagnetic fields. Various electromagnetic levitation schemes for transport (viz. Maglev scheme, etc.) are examples of such energy combinations. It should also be noted that in a number of electromagnetic levitation devices, the constraints for the stability are themselves not electromagnetic in nature, though there are systems where both the suspension as well as the stability have been achieved by electromagnetic means, such as Laithwaite's 'Magnetic River Principle'.

PROBLEMS

11.1 Two coils having the same self-inductance are connected in series. When a current I flows through the coils, the magnetic energy stored in their fields is W joules. If the connections of one coil are interchanged and the current reduced to $I/2$, the energy stored again is W joules. Calculate the ratio of the mutual inductance to the self-inductance.

Ans.: $M/L = 3/5$

11.2 A current I flows round a circle of radius a and a current I' flows in a very long wire in the same plane. Show that the mutual attraction is $\mu_0 II'(\sec \alpha - 1)$ where 2α is the angle subtended by the circle at the nearest point of the straight wire.

11.3 Two long solenoids, one of which extends a distance l within the other, have currents flowing in the same direction in them. Neglect the end effects and assume that their diameters are approximately equal (cross-sectional area being A) for simplicity of calculations. Show that the force between them is axial and attractive (why?) and is given by

$$F = \mu_0 N_a N_b I_a I_b A$$

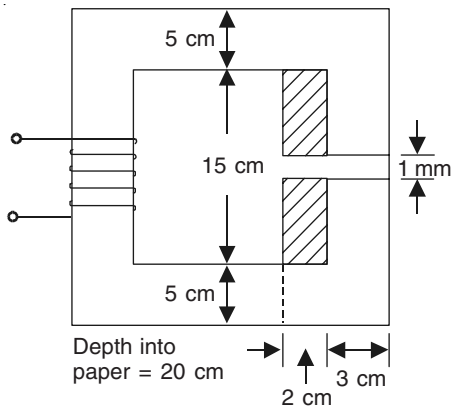
where N_a and N_b are respectively the turns on the two solenoids per metre, carrying currents I_a and I_b respectively. Solve the problem by both the methods, i.e. (i) evaluating the mutual inductance, and (b) considering the magnetic energy of the system.

11.4 The diagram shows an air-cored choke, having 200 turns wound on a laminated core of iron. Estimate the inductance:

(i) when the magnetic circuit is as shown by full lines

(ii) when the portion hatched is removed.

In both the cases, assume that the iron is infinitely permeable, and neglect leakage and fringing. In practice, the iron can be considered to be completely saturated at $B = 1.8T$. Show that this means the choke can be used satisfactorily for currents up to approximately 7 amps.



Ans.: (i) $L_1 = 503$ mH (ii) $L_2 = 302$ mH

11.5 An electromagnet has opposed poles having square faces of side l, separated by an air-gap G. Into this air-gap is moved an iron plate with faces parallel to the pole faces and its edges parallel to the pole-edges; its length and width are l, but its thickness is $(G - g)$. The plate overlaps the poles by a distance x in one direction, and in the perpendicular direction they overlap completely. The remaining magnetic path of the electromagnet is of iron and U-shaped, and is wound with a winding of N turns carrying a current I; but the exact shape is immaterial, since the iron of both the magnet and the plate is of zero reluctance so that the mmf NI is solely employed in forcing the flux across the air-gaps. Prove the following:

(i) The flux traversing the air-gap is $\mu_0 NIl\left(\frac{l-x}{G}+\frac{x}{g}\right)$

(ii) The energy stored in the air-gap field is $=\frac{1}{2}\mu_0 N^2 I^2 l\left(\frac{l-x}{G}+\frac{x}{g}\right)$

(iii) The force tending to draw the iron plate further into the air-gap is

$$=\frac{1}{2}\mu_0 N^2 I^2 l\left(\frac{G-g}{Gg}\right)$$

Flux fringing should be neglected.

11.6 Find the cyclotron frequency for electrons in a magnetic flux density of 0.2 Wb/m^2.

Ans.: 5600 MHz

11.7 An ultraviolet radiation releases photo-electrons from a plate at negligible initial velocity. A magnetic field **B** is maintained parallel to the plate and the **E** field normal to it. The **E** field is generated by using a second plate kept parallel to the first plate, at a distance d from it and at a potential $+V$ with respect to it. Prove that the value of d for which current just fails to pass between the plates is $\sqrt{(2mV/eB^2)}$, where e = charge of the electron, and m = mass of the electron.

12 Maxwell's Equations

12.1 INTRODUCTION

So far we have studied the special cases of the electromagnetic fields. We started with the electric fields produced by the static electric charges, then proceeded to the electric currents produced by the moving charges, the magnetic fields due to the steady currents, and the electromagnetic fields due to the quasi-stationary (i.e. low-frequency alternating) currents. The method of analysis which we have used so far has been inductive. Our starting point was the Coulomb's law of force between particle charges, slowly moving in an arbitrary manner relative to the observer, and a model of the atomic structure of matter. From there we defined the concepts of the electric and the magnetic fields and investigated the properties of these vector fields. This approach has given us a deep insight into the basic relations between the electric and the magnetic fields, and thus enabled us to understand various related problems in engineering with elegance and simplicity.

However a stage has now been reached, where our knowledge is sufficient to let us consider the most general types of the electromagnetic fields. We shall see now that, with some further modifications or additions to the relations which have been derived for the special fields, a consistent system of generally valid equations is obtained. These equations are known as the 'Maxwell's electromagnetic field equations'. They are, in fact, not new for us. We have already come across them in various places during our present study, under different restricted conditions.

Maxwell's equations, when expressed in differential form, comprise a system of partial differential equations, which the electric and the magnetic components of the unique electromagnetic field must satisfy in all cases. All which have been derived earlier, are special cases of these general equations. Based on the work done by his predecessors like Michael Faraday, Ampere, and others, Maxwell unified the relationships between the electric and the magnetic fields, and expressed the generalized unification in four equations written down in integral form. As mentioned earlier, some of these equations had been derived in restricted form by the previous workers in these fields. However it was Maxwell's insight into these apparently disparate areas that enabled him to combine and extrapolate them to achieve a unified picture. These unified equations, in fact, anticipated the possibility of existence of the varying electromagnetic fields independent of the 'conduction currents'. These predictions were proved experimentally some twenty years later by Heinrich Hertz, a German army officer. The propagation of energy in the

form of the time-varying electromagnetic fields along the transmission lines, and in the wave-guides, as well as the behaviour of the antennae cannot be analyzed without an understanding of the Maxwell's equations.

In this chapter, we shall introduce these equations, first in the integral form and then in the more familiar differential form along with the associated relationships, and analyze some of their general important consequences. In the later chapters, the analysis will start from the Maxwell's equations, i.e. the method of analysis will no longer be inductive, but rather be deductive. Special cases, ranging from the low-frequency eddy currents to high frequency wave propagation, will be analyzed by means of these general equations.

The reader must take great care to read and to understand the contents of this chapter carefully.

12.2 THE EQUATION OF CONTINUITY (OR THE LAW OF CONSERVATION OF ELECTRIC CHARGE)

We have already discussed in detail the equation of continuity in the Section 6.4. However we shall very briefly recapitulate the mathematical steps in its derivation, as this is an essential step in explaining the generalization which would lead us to the Maxwell's equations.

We consider a fixed closed surface Σ in a region containing a space charge of density ρ_C which is changing with time. This means that there must be currents, specified by a current density vector $\mathbf{J}$.

$\therefore$ The rate of loss of charge from within Σ is given by

$$\oiint_S \mathbf{J} \cdot d\mathbf{S} = \iiint_v (\text{div } \mathbf{J})\, dv \tag{12.1}$$

where v is the volume enclosed by Σ (Figure 12.1). The above equation is a consequence of the 'divergence theorem.'

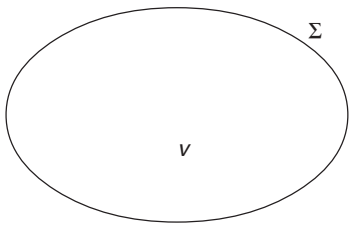


Figure 12.1 A closed surface with space charge of density ρ_C.

At an instant, the charge remaining in the volume v is given by

$$\iiint_v \rho_C\, dv$$

$\therefore$ The rate of loss of charge $= -\left(\frac{\partial}{\partial t}\right) \iiint_v \rho_C\, dv$ (12.2)

∴ Equating these two rates of loss of charge,

$$\oiint_{\Sigma} \mathbf{J} \cdot d\mathbf{S} = -\left(\frac{\partial}{\partial t}\right) \iiint_{v} \rho_C \, dv$$

Or, substituting from Eq. (12.1),

$$\text{div } \mathbf{J} + \frac{\partial \rho_C}{\partial t} = 0 \tag{12.3}$$

This is the equation of continuity, expressing the fact that the electric charges can neither be created nor be destroyed in macroscopic quantities.

12.2.1 A Note on the Continuity Equation

Just now, we have seen that the equation of continuity is

$$\text{div } \mathbf{J} + \frac{\partial \rho_C}{\partial t} = 0 \tag{12.3}$$

Gauss' theorem in electrostatics states that

$$\text{div } \mathbf{D} = \nabla \cdot \mathbf{D} = \rho_C$$

Substituting in Eq. (12.3), we get

$$\text{div } \mathbf{J} + \frac{\partial}{\partial t} (\text{div } \mathbf{D}) = 0$$

Since the space operator div is independent of the time-operator $(\partial/\partial t)$, we can reverse their order in this equation.

$$\therefore \quad \text{div } \mathbf{J} + \text{div}\left(\frac{\partial \mathbf{D}}{\partial t}\right) = 0$$

or

$$\text{div}\left(\mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}\right) = 0 \tag{12.4}$$

i.e. $\mathbf{J} + \dfrac{\partial \mathbf{D}}{\partial t}$ is a solenoidal vector.

This vector will be found to be very important, e.g. we consider the discharge of a capacitor through a resistive circuit (Figure 12.2). If one considers a closed surface like Σ_1 on a part of this circuit (as shown in the figure), then the current leaving Σ_1 equals the current entering it. The

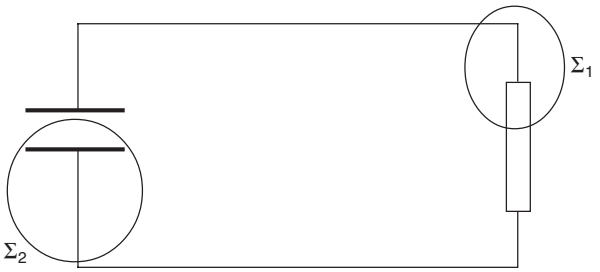


Figure 12.2 Discharge of a capacitor through a resistor.

same cannot be said of the closed surface Σ_2, until one admits a displacement current ($\partial \mathbf{D}/\partial t$) as equivalent to the conduction current $\mathbf{J}$. But with this equivalence, we have a closed circuit, which takes the form of a conduction current in the wire, and a displacement current in the capacitor dielectric. Equation (12.4) underlies the treatment of the capacitor as a circuit element.

12.3 DISPLACEMENT CURRENT AND THE GENERALIZATION OF THE MAGNETIC CIRCUIT LAW

If the Ampere's magnetic circuit law in its elementary form is applied to a small circuit C_x, of area $\delta \mathbf{S}_x$ directed along the coordinate axis Ox as shown in Figure 12.3, we get

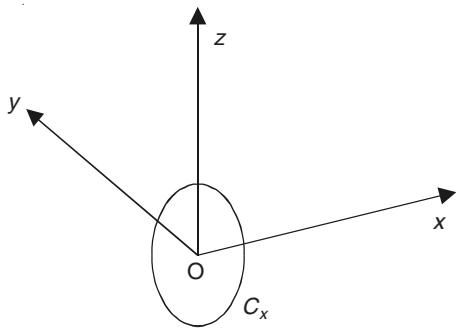


Figure 12.3 Loop for Ampere's law.

$$\oint_{C_x} H_1 \, dl = J_x \, \delta S_x \tag{12.5}$$

But the left-hand side of this equation = $\delta S_x \, \text{curl}_x \, \mathbf{H}$.

Thus Eq. (12.5) reduces to

$$\text{curl}_x \, \mathbf{H} = J_x$$

Or, taking account of the other two components, we get

$$\text{curl} \, \mathbf{H} = \nabla \times \mathbf{H} = \mathbf{J} \tag{12.6}$$

but this equation cannot be universally true; for we have seen that for any arbitrary vector $\mathbf{A}$, div curl $\mathbf{A} = 0$ (an identity); hence the truth of Eq. (12.6) must depend on

$$\text{div} \, \mathbf{J} = 0$$

We have seen however from the continuity equation,

$$\text{div} \, \mathbf{J} = -\frac{\partial \rho_C}{\partial t}$$

so that the condition div $\mathbf{J} = 0$ is not satisfied.

On the other hand, we did discover that

$$\text{div}\left(\mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \right) = 0$$

Thus Maxwell postulated that the current density in the magnetic circuit law is not **J** but [**J** + (∂**D**/∂t)], and he called (∂**D**/∂t) the 'displacement current density', and [**J** + (∂**D**/∂t)] is called the 'total current density'.

Working out the consequences of this, Maxwell proved that the electromagnetic signals could be transmitted with a finite velocity equal to the measured velocity of light in free space. This led him to the 'Electromagnetic Theory of Light', and mankind to radio, telegraphy, etc. Maxwell's postulate therefore accords with the observed results, and the general form of the magnetic circuit law is proved to be

$$\text{curl } \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \tag{12.7}$$

where **J** is the conduction current density, and (∂**D**/∂t) is the displacement current density.

It will be seen that the displacement current can be neglected, in general, at low frequencies. For example, considering a conductor of resistivity ρ (say), we have

$$\mathbf{D} = \varepsilon_0 \varepsilon_r \mathbf{E} \qquad \text{and} \qquad \mathbf{E} = \rho \mathbf{J}$$

$$\therefore \qquad \mathbf{D} = \varepsilon_0 \varepsilon_r \rho \mathbf{J}, \quad \frac{\partial \mathbf{D}}{\partial t} = \varepsilon_0 \varepsilon_r \rho \left(\frac{\partial \mathbf{J}}{\partial t} \right)$$

Let $\mathbf{J} = \hat{\mathbf{J}} \exp(j\omega t)$ (i.e. sinusoidally varying with time).

$\therefore$ The ratio of displacement current to the conduction current

$$= \frac{\dfrac{\partial \mathbf{D}}{\partial t}}{\mathbf{J}} = \varepsilon_0 \varepsilon_r \rho \omega$$

For a conducting metal like copper,

$$\varepsilon_r = 1, \qquad \rho = 1.7 \times 10^{-8} \ \Omega\text{-m}, \qquad \varepsilon_0 = 8.854 \times 10^{-12}$$

$\therefore$ The ratio $= 8.854 \times 10^{-12} \times 1 \times 1.7 \times 10^{-8} \times 2\pi f$

$= 9.46 \times 10^{-19} f \approx 10^{-18} f$

$\therefore$ The displacement current is negligible in the total current at the normal power frequency of 50 Hz, and easily up to the radio frequencies.

12.4 MAXWELL'S EQUATIONS AND THE CONSTITUTIVE RELATIONS

We shall now state the Maxwell's equations, using the electric and the magnetic field vectors which we have studied so far and generalized them, i.e. **E**, **H**, **D**, and **B**, etc. to make them all inclusive taking account of all kinds of different media. The equations will be written in terms of all these generalized vectors, though when we refer to a specific medium or to a restricted condition, we shall explicitly state the restrictions and draw suitable conclusions.

Maxwell's equations in integral forms will be stated first, even though these equations in differential form are more widely used. The reason for this approach is that the integral forms of these equations are more fundamental in nature than the differential forms, for the reason that these equations are used for deriving the boundary conditions on surfaces or lines of discontinuity. These equations in differential forms require implicitly the assumption of continuity and

hence cannot be used for deriving such conditions. However since the differential forms of these equations have been derived from the integral forms, it should be understood that the continuity requirement is not a necessary condition for the satisfaction of these equations. Hence the mathematically rigorous approach to study these equations is to start from their integral forms.

12.4.1 The Integral Form of Maxwell's Equations

The four basic equations derived so far are collectively known as the Maxwell's equations. The first two concern the flux out of any arbitrarily chosen closed surface *S*, and the last two concern the line-integral round any arbitrarily chosen closed contour or circuit *C*. They are:

$$\oiint_S \mathbf{D} \cdot d\mathbf{S} = \iiint_v \rho_C dv \tag{12.8}$$

(also known as the Gauss' theorem)

and

$$\oiint_S \mathbf{B} \cdot d\mathbf{S} = 0 \tag{12.9}$$

(for the solenoidal property of **B**)

The next two equations are

$$\oint_C \mathbf{E} \cdot d\mathbf{l} = -\left(\frac{d}{dt}\right) \iint_S \mathbf{B} \cdot d\mathbf{S} \tag{12.10}$$

(generalized Faraday's law of Induction)

and

$$\oint_C \mathbf{H} \cdot d\mathbf{l} = \iint_S \left(\mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}\right) \cdot d\mathbf{S} \tag{12.11}$$

(generalized Ampere's Magnetic Circuit law)

The equation of continuity is implicit in the last equation (12.11), since it can take the form

$$\text{div}\left(\mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}\right) = 0$$

These are the four equations with six unknowns, and hence need supplementing in any given problem by others which are called the 'constitutive equations', because they describe, perhaps approximately the constitution of the materials in the field.

In free space, or in dielectric, the relation between **D** and **E** is

$$\mathbf{D} = \varepsilon_0 \varepsilon_r \mathbf{E} = \varepsilon \mathbf{E} \tag{12.12}$$

which is accurate for vacuum, and a good approximation for the behaviour of many homogeneous dielectrics.

In free space, or any material, the relation between **B** and **H** is given by

$$\mathbf{B} = \mu_0 \mu_r \mathbf{H} = \mu \mathbf{H} \tag{12.13}$$

which is accurate for vacuum, and for non-magnetic materials. For ferromagnetic materials, μ_r is not a constant.

Lastly, for ohmic conductors,

$$\mathbf{E} = \rho \mathbf{J}$$

or

$$\mathbf{J} = \sigma \mathbf{E} \tag{12.14}$$

which is a good approximation for the behaviour of the metals. So we can say that, in general, these constants, i.e. ε_r (relative permittivity) and μ_r (relative permeability) (and ρ or σ for the metals) are only good approximations for most media, except for the free space, where they are constants rigorously. However for the progress of our analysis, the assumption of linearity is essential.

The physical interpretations of the four equations of Maxwell can be stated briefly as given below:

The first equation [Eq. (12.8)] is the extension of Coulomb's law for point charges to distributed charges as well as to charge clouds. This is also known as Gauss' theorem.

The second equation [Eq. (12.9)] is a mathematical statement of the fact that magnetic poles (i.e. monopoles) do not exist in isolation in nature.

The third equation [Eq. (12.10)] is the mathematical form of the generalized Faraday's law of electromagnetic induction which has been discussed in detail earlier in Chapter 10.

The fourth equation [Eq. (12.11)] is really the completion and extrapolation of Ampere's law to take account of the displacement current (i.e. $\frac{\partial \mathbf{D}}{\partial t}$) from which Maxwell theoretically predicted the existence of electromagnetic waves (refer to Sections 12.5 and 12.6).

12.4.2 Maxwell's Equations in Differential Form

Rewriting the four Maxwell's equations, i.e. (12.8) to (12.11) in differential form, we have

$$\text{div } \mathbf{D} = \nabla \cdot \mathbf{D} = \rho_C \tag{12.15}$$

$$\text{div } \mathbf{B} = \nabla \cdot \mathbf{B} = 0 \tag{12.16}$$

$$\text{curl } \mathbf{E} = \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{12.17}$$

$$\text{curl } \mathbf{H} = \nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \tag{12.18}$$

As before, the equation of continuity is a consequence of these equations. These equations, together with the constitutive relations mentioned in Section 12.4.1, comprise the complete set of the equations.

12.4.3 Complex Representation of Maxwell's Equations

When the time-variation of all the field vectors is sinusoidal (i.e. the time-harmonic type), the vectors can be written in the complex variable form [i.e. $\mathbf{A}$ exp $(j\omega t)$ type], and then these equations become (from Sections 0.9 to 0.93)

$$\nabla \cdot \mathbf{D}_c = \rho_C \tag{12.19}$$

$$\nabla \cdot \mathbf{B}_c = 0 \tag{12.20}$$

$$\nabla \times \mathbf{E}_c = -j\omega \mathbf{B}_c \tag{12.21}$$

$$\nabla \times \mathbf{H}_c = \mathbf{J}_c + j\omega \mathbf{D}_c \tag{12.22}$$

and the equation of continuity is now

$$\nabla \cdot \mathbf{J}_c + j\omega \rho_C = 0 \tag{12.23}$$

where the suffix c for the vectors stands for the complex forms.

12.4.4 Historical Comments on Maxwell's Equations

In fact the 'Maxwell's equations' as have been presented here were not how Maxwell wrote them. Though Maxwell thought vectorially (as shown by his invention of curl, div and grad), his results had been presented in triple component Cartesian form, and the compact vector notation had to wait for the work of Heaviside and Willard Gibbs. Maxwell wrote the equations using 'Quaternions' as discovered by the Irish mathematician, William Rowan Hamilton. This was basically a combination of a scalar potential and a vector quantity somewhat akin to the vector potential. In particular the equation for Faraday's law of electromagnetic induction was written down by Maxwell in terms of the magnetic vector potential (?) **A** which he named 'electrokinetic momentum vector'. This name itself would give some idea about the physical interpretation of this device. At that time 'vector calculus' had not been discovered, and it was subsequently discovered by the self-taught British physicist and engineer Oliver Heaviside and the American mathematician J. Willard Gibbs of Princeton University. A point that should be noted in this context is that the reason for the discovery of vector calculus was 'problems in electrodynamics and not any problems in applied mechanics' as might seem to be more likely.

Maxwell was a person far ahead of his times and it took scientists another 20 years or so before they could verify experimentally what Maxwell had predicted from his theoretical analysis. The present form of these equations as stated here (the most familiar version) which has a type of symmetry in terms of the four vectors, i.e. **D**, **E**, **H** and **B**, was really given by the eccentric British genius Oliver Heaviside and used by the German physicist Heinrich Hertz for the experimental verification of Maxwell's prediction. So initially they were called

Heaviside–Hertz form of Maxwell's equations.

Heaviside considered Maxwell's analysis as the work of a "heaven-born genius".

12.5 CONSEQUENCES OF MAXWELL'S EQUATIONS

The four equations as postulated by Maxwell have been expressed in terms of the five field vectors **E**, **D**, **J**, **H**, and **B**, which are interrelated between themselves by the three constitutive relations. Since these equations, particularly the 'curl' equations are in terms of more than one field vector, we shall now reduce these equations to a form in which the operating equation is expressed as a function of single field vector. We shall start with the differential form of these equations, i.e. (12.15) to (12.18), and the constitutive relations (12.12) to (12.14). We will also assume a region of linear media, i.e. the characteristic constants ε_r, μ_r, and σ (or ρ) as constants, and the time-variation as sinusoidal which will permit us to use the complex representation of the field vectors. However for the sake of simplicity and generality, we shall not use the complex suffix notation c with the field vectors. So we start with Eq. (12.18), which is

$$\text{curl } \mathbf{H} = \nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} = \sigma \mathbf{E} + \varepsilon \left(\frac{\partial \mathbf{E}}{\partial t} \right) \tag{12.24}$$

using the constitutive relations.

Taking the curl operation again on this equation, we have

$$\nabla \times \nabla \times \mathbf{H} = \sigma(\nabla \times \mathbf{E}) + \varepsilon \left[\nabla \times \left(\frac{\partial \mathbf{E}}{\partial t}\right)\right] = \sigma(\nabla \times \mathbf{E}) + \varepsilon \left(\frac{\partial}{\partial t}\right)(\nabla \times \mathbf{E})$$

$$= -\sigma\left(\frac{\partial \mathbf{B}}{\partial t}\right) + \varepsilon\left(\frac{\partial}{\partial t}\right)\left(-\frac{\partial \mathbf{B}}{\partial t}\right)$$

$$= -\mu\sigma\left(\frac{\partial \mathbf{H}}{\partial t}\right) - \mu\varepsilon\left(\frac{\partial^2 \mathbf{H}}{\partial t^2}\right) \tag{12.25}$$

using Eq. (12.17) and the constitutive relations.

We now consider the vector identity:

$$\nabla \times \nabla \times \mathbf{H} = \nabla\,(\nabla \cdot \mathbf{H}) - \nabla^2 \mathbf{H} = \text{grad (div } \mathbf{H}) - \nabla^2 \mathbf{H} \tag{12.26}$$

Since we are considering linear media at present, from Eq. (12.16), we have

$$\text{div } \mathbf{B} = \text{div } \mu\mathbf{H} = \mu(\text{div } \mathbf{H}) = 0 \tag{12.27}$$

and it should be noted that the operator ∇^2 is the Laplacian operator, which (say) in Cartesian coordinates is

$$\nabla^2 \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{12.28}$$

and when it is operated on a vector, the result is a compact expression for this operator operating on all the three components of the vector, e.g. in Cartesian coordinates:

$$\nabla^2 \mathbf{H} = \mathbf{i}_x \nabla^2 H_x + \mathbf{i}_y \nabla^2 H_y + \mathbf{i}_z \nabla^2 H_z \tag{12.29}$$

$\therefore$ From Eqs. (12.26) and (12.27), we get

$$\nabla \times \nabla \times \mathbf{H} = -\nabla^2 \mathbf{H} \tag{12.30}$$

Substituting Eq. (12.30) in Eq. (12.25),

$$\nabla^2 \mathbf{H} = \mu\sigma\left(\frac{\partial \mathbf{H}}{\partial t}\right) + \mu\varepsilon\left(\frac{\partial^2 \mathbf{H}}{\partial t^2}\right) \tag{12.31}$$

and using the constitutive relation (12.13) in the linear media, from the above equation we get

$$\nabla^2 \mathbf{B} = \mu\sigma\left(\frac{\partial \mathbf{B}}{\partial t}\right) + \mu\varepsilon\left(\frac{\partial^2 \mathbf{B}}{\partial t^2}\right) \tag{12.32}$$

Thus the magnetic field vectors both satisfy the same operating equation.

Next, we consider Eq. (12.17), and take a second curl operation on it as before, and we get

$$\nabla \times \nabla \times \mathbf{E} = -\nabla \times \left(\frac{\partial \mathbf{B}}{\partial t}\right) = -\mu\left(\frac{\partial}{\partial t}\right)(\nabla \times \mathbf{H})$$

$$= -\mu\left(\frac{\partial}{\partial t}\right)\left(\mathbf{J}+\frac{\partial \mathbf{D}}{\partial t}\right) = -\mu\left(\frac{\partial}{\partial t}\right)\left[\sigma\mathbf{E}+\varepsilon\left(\frac{\partial \mathbf{E}}{\partial t}\right)\right]$$

$$= -\mu\sigma\left(\frac{\partial \mathbf{E}}{\partial t}\right)-\mu\varepsilon\left(\frac{\partial^2 \mathbf{E}}{\partial t^2}\right) \tag{12.33}$$

[this is similar to Eq. (12.25)], and

$$\nabla \times \nabla \times \mathbf{E} = \nabla\,(\nabla \cdot \mathbf{E}) - \nabla^2\mathbf{E} = \text{grad (div } \mathbf{E}) - \nabla^2\mathbf{E} \tag{12.34}$$

Gauss' theorem gives [i.e. Eq. (12.15)] us:

$$\text{div } \mathbf{D} = (\nabla \cdot \mathbf{D}) = \rho_C \leftarrow \text{the charge density}$$

$\therefore$ In the charge-free region *only*, $\rho_C = 0$.

$\therefore$ $(\nabla \cdot \mathbf{D}) = 0$ and $(\nabla \cdot \mathbf{E}) = 0$.

$\therefore$ Combining this with Eqs. (12.33) and (12.34), we get

$$\nabla^2\mathbf{E} = \mu\sigma\left(\frac{\partial \mathbf{E}}{\partial t}\right)+\mu\varepsilon\left(\frac{\partial^2 \mathbf{E}}{\partial t^2}\right) \tag{12.35}$$

Since the vectors **D** and **J** are related to **E** by the constant factors in any linear media, both **D** and **J** also satisfy the same equation.

$\therefore$ In the charge-free and linear region, all the five field vectors satisfy the same operating equation of the form derived in the relation (12.31) or (12.32) or (12.35).

We shall now consider the time-varying fields, and consider the type of variation most commonly used, i.e. the steady-state alternating current operation of the form given by (Sections 0.9–0.9.1)

$$\mathbf{H} = \hat{\mathbf{H}} \exp(j\omega t) \qquad \text{(i.e. sinusoidal variation with time)}$$

The time-derivatives in these equations then become

$$\frac{\partial}{\partial t} = j\omega \qquad \text{and} \qquad \frac{\partial^2}{\partial t^2} = -\omega^2,\ \omega \text{ being the angular frequency} = 2\pi f, \text{ and } f = \text{frequency}$$

And the Eqs. (12.31) and (12.35) become

$$\nabla^2\mathbf{H} = (j\omega\mu\sigma - \omega^2\mu\varepsilon)\mathbf{H}$$

$$\nabla^2\mathbf{E} = (j\omega\mu\sigma - \omega^2\mu\varepsilon)\mathbf{E} \tag{12.36}$$

In fact, all the five field vectors satisfy the same operating equation. These equations are known as the 'Vector Helmholtz equations'. It will be seen that $1/\sqrt{\mu\varepsilon}$ has the dimension of velocity, and in free space $\mu\varepsilon = \mu_0\varepsilon_0 = 1/c^2$, where c is the velocity of propagation of light, e.g.

$$\mu_0\varepsilon_0 = (4\pi \times 10^{-7}) \text{ henries/metre} \times [10^{-9}/36\pi] \text{ farads/metre}$$

$$= [10^{-16}/9]\ [\text{coulombs}^2/\text{newton-metres}^2 \times \text{newtons/ampere}^2]$$

and ampere = coulombs/second

$$= [1/(9 \times 10^{16})]\ [(\text{second/metre})^2]$$

$$\therefore \quad c = 1/\sqrt{\mu_0 \varepsilon_0} = (3 \times 10^8) \text{ metres/second}$$

And, we substitute $(\omega/c) = \beta$, and Eqs. (12.36) become

$$\nabla^2 \mathbf{H} = (j\omega\mu\sigma - \beta^2)\mathbf{H}$$

$$\nabla^2 \mathbf{E} = (j\omega\mu\sigma - \beta^2)\mathbf{E} \tag{12.37}$$

Limiting cases

1. If the conductivity of the medium is very low (i.e. σ is very low, or ρ is very high), then $\beta^2 >> \omega\mu\sigma$, and in the above equations the $(\omega\mu\sigma)$ term can be neglected. Hence Eqs. (12.37) simplify to:

$$\nabla^2 \mathbf{H} + \beta^2 \mathbf{H} = 0$$

$$\nabla^2 \mathbf{E} + \beta^2 \mathbf{E} = 0 \tag{12.38}$$

This is known as the vector wave equation. It considers only the displacement current, and is used for the wave propagation, waveguides, and plasmas, etc. problems.

2. When the conductivity σ is very high, such that $\beta^2 << \omega\mu\sigma$, then the β^2 term becomes negligible, and Eqs. (12.37) simplify to:

$$\nabla^2 \mathbf{H} - j\omega\mu\sigma\mathbf{H} = 0$$

$$\nabla^2 \mathbf{E} - j\omega\mu\sigma\mathbf{E} = 0 \tag{12.39}$$

This is the vector diffusion-type equation, and considers the effects of the conduction currents only. It is used in all skin effect and eddy current problems (in metallic conductors).

12.6 PLANE ELECTROMAGNETIC WAVES IN FREE SPACE

We shall now consider the Maxwell's equations to investigate the simplest pattern of electromagnetic fields in free space, independently (at this stage) of the means by which the fields are produced. In free space, the charge density $\rho_C = 0$, and, the conduction current density $\mathbf{J} = 0$; and also, $\mathbf{D} = \varepsilon_0\mathbf{E}$, and $\mathbf{B} = \mu_0\mathbf{H}$.

We use the right-handed rectangular Cartesian coordinate system. Let the electric field vector **E** have one component only E_x (hereafter to be called E), and let this vary with z and time t only, i.e. the value is uniform at any instant over the whole of any plane perpendicular to Oz (Figure 12.4). [In other words, the coordinates are so chosen (for a uniform plane wave, to be defined later) as to simplify the mathematical analysis without any loss of generality]. We could have started from the general wave equation which we have derived before, but for clarity, we will start from the Maxwell's equations, simplify each for the present problem, and derive the one-dimensional wave equation for the plane wave as defined by the vector **E** field which has no variation normal to the z-direction; i.e.

$$E_x = E_x(z, t) = E(z, t) = E, \qquad E_y = 0, \qquad E_z = 0$$

The components of curl **E** are then:

$$\text{curl } \mathbf{E} = \nabla \times \mathbf{E}$$

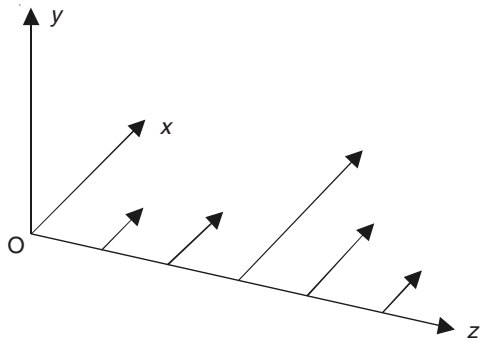


Figure 12.4 A uniform plane wave.

$$= \mathbf{i}_x \left(\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} \right) + \mathbf{i}_y \left(\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} \right) + \mathbf{i}_z \left(\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} \right)$$

$$= \mathbf{i}_x 0 + \mathbf{i}_y \left(\frac{\partial E_x}{\partial z} \right) + \mathbf{i}_z 0$$

$\therefore$ Maxwell's equation: curl $\mathbf{E} = -\left(\dfrac{\partial \mathbf{B}}{\partial t} \right)$ gives:

$$\left(\frac{\partial B_x}{\partial t} \right) = 0, \qquad \left(\frac{\partial B_y}{\partial t} \right) = -\left(\frac{\partial E_x}{\partial z} \right), \qquad \left(\frac{\partial B_z}{\partial t} \right) = 0$$

Setting aside any possible direct (steady) component like earth's magnetic field, this shows that the time-varying part of **B** (associated with the defined **E** field) has only the *y*-component (to be called from now on as *B*), related to *E* (which is E_x) by the equation:

$$\frac{\partial E}{\partial z} = -\frac{\partial B}{\partial t} \tag{12.40}$$

In absence of **J**, the other Maxwell's equation is

$$\text{curl } \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t}$$

Also in free space,

$$\mathbf{H} = \frac{\mathbf{B}}{\mu_0} \qquad \text{and} \qquad \mathbf{D} = \varepsilon_0 \mathbf{E}$$

Hence,

$$\text{curl } \mathbf{B} = \mu_0 \varepsilon_0 \left(\frac{\partial \mathbf{E}}{\partial t} \right)$$

and with **E** and **B** restricted to the components described, this becomes

$$\frac{\partial B}{\partial z} = -\mu_0 \varepsilon_0 \left(\frac{\partial E}{\partial t} \right), \qquad \left[\text{which is, } -\frac{\partial B_y}{\partial z} = \mu_0 \varepsilon_0 \left(\frac{\partial E_x}{\partial t} \right) \right] \tag{12.41}$$

Differentiating Eq. (12.40) with respect to *z* and Eq. (12.41) with respect to *t*, we get

$$\frac{\partial^2 E}{\partial z^2} = -\frac{\partial^2 B}{\partial z \partial t} = +\mu_0 \varepsilon_0 \left(\frac{\partial^2 E}{\partial t^2} \right)$$

We write $\mu_0 \varepsilon_0 = \dfrac{1}{c^2}$, and so get

$$\frac{\partial^2 E}{\partial z^2} = \left(\frac{1}{c^2} \right) \left(\frac{\partial^2 E}{\partial t^2} \right) \tag{12.42}$$

This is the one-dimensional wave equation; and it is satisfied by

$$E = f(z \pm ct) \tag{12.43}$$

where f is any function. For

$$\frac{\partial E}{\partial z} = \pm f'(z \pm ct) \quad \text{and} \quad \frac{\partial^2 E}{\partial z^2} = f''(z \pm ct)$$

where f' and f'' are derivatives with respect to the whole argument. Similarly

$$\frac{\partial E}{\partial t} = \pm c f'(z \pm ct) \quad \text{and} \quad \frac{\partial^2 E}{\partial t^2} = c^2 f''(z \pm ct)$$

Therefore Eq. (12.42) is satisfied by Eq. (12.43).

We consider first the solution,

$$E = f(z - ct)$$

and we have from Eq. (12.41)

$$\frac{\partial B}{\partial z} = -\left(\frac{1}{c^2} \right) \left(\frac{\partial E}{\partial t} \right) = +\left(\frac{1}{c} \right) f'(z - ct)$$

whence

$$B = \left(\frac{1}{c} \right) f(z - ct) = \frac{E}{c} \tag{12.44}$$

Thus E and B are proportional. If we compare their values at times $t = 0$ and $t = t_1$ (see Figure 12.5), we see that the shape of the pattern is unchanged, but the whole pattern has moved to the right by a distance equal to ct_1. We call this moving pattern a travelling wave. The velocity is c, given by

$$c = (\mu_0 \varepsilon_0)^{-1/2} = 2.998 \times 10^8 \text{ m/s} \tag{12.44a}$$

This equals the measured velocity of light in free space, and it led Maxwell to identify light as an electromagnetic radiation. To him, c came as the ratio of the electromagnetic to the electrostatic unit of charge.

In a similar way, we shall see that $E = f(z + ct)$ describes a wave pattern travelling to the left; and in this pattern $B = -E/c$. Note also that the direction of propagation of the wave (labelled $\rightarrow\rightarrow$ in Figure 12.6 below) is that of the vector product ($\mathbf{E} \times \mathbf{B}$) or since $\mathbf{B} = \mu_0 \mathbf{H}$, of ($\mathbf{E} \times \mathbf{H}$). This later will prove to be highly significant.

Since

$$B = \frac{E}{c} = \left(\sqrt{\mu_0 \varepsilon_0} \right) E$$

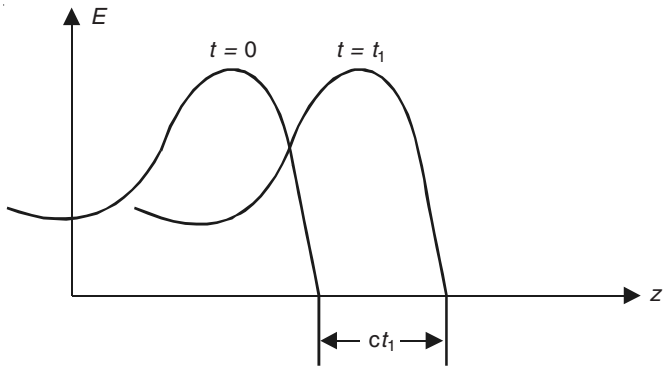


Figure 12.5 A travelling wave pattern.

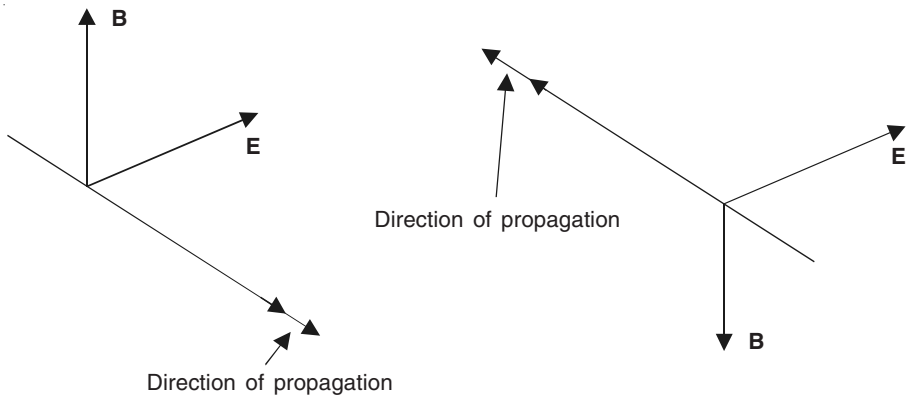


Figure 12.6 Direction of propagation of **E** and **B** (or **H**) waves.

$$H = \frac{B}{\mu_0} = \left(\sqrt{\frac{\varepsilon_0}{\mu_0}}\right)E \qquad \text{or} \qquad E = \left(\sqrt{\frac{\mu_0}{\varepsilon_0}}\right)H \tag{12.45}$$

Now E is in volts/metre and H is in amps/metre. Hence $\sqrt{\mu_0/\varepsilon_0}$ is measured in ohms, and we write

$$Z_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}} = 376.7\ \Omega \tag{12.46}$$

Z_0 is called the *Wave Impedance* or the *Intrinsic Impedance of free space*.

The wave system described here is said to be a *plane* wave, because **E** and **B** are constant over planes perpendicular to the direction of propagation Oz.

The most important case is that of the waves in which **E** and **B** are sinusoidal functions of $(z \pm ct)$. The solution to the **E** field may then be written as

$$E = \hat{E} \cos K(z - ct), \text{ for the forward travelling wave.}$$

If ω is the angular frequency, then $Kc = \omega$. Hence substituting for K, we get

$$E = \hat{E} \cos \omega\left(t - \frac{z}{c}\right)$$

and

$$B = \hat{B}\cos\omega\left(t - \frac{z}{\mathrm{c}}\right)$$

where

$$\hat{B} = \frac{\hat{E}}{\mathrm{c}} \tag{12.47}$$

There is thus a phase lag of $\omega z/\mathrm{c}$ in the oscillations at $z = z$ compared with $z = 0$; planes perpendicular to Oz are planes of constant phase, and so c is now referred to as the *phase velocity*. The wave length λ of the wave is the distance corresponding to a phase change of 2π; thus

$$2\pi = \frac{\omega\lambda}{\mathrm{c}} \qquad \text{or} \qquad \omega\lambda = 2\pi\mathrm{c} \qquad \text{or} \qquad f\lambda = \mathrm{c} \tag{12.48}$$

where f is the frequency of oscillation of the wave, $2\pi f = \omega$ (Figure 12.7).

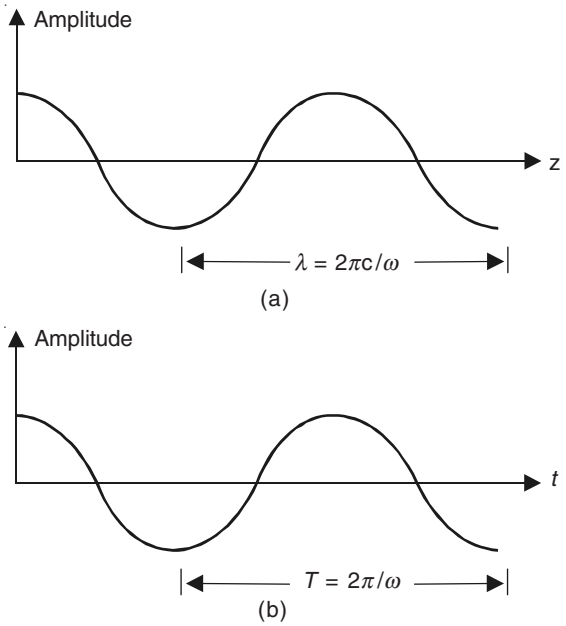


Figure 12.7 (a) Wavelength and (b) time period of a sinusoidal wave.

The quantity ω/c (or more generally ω/u, where u = velocity) is called the *wave number* and is denoted as k or β, i.e.

$$k = \frac{\omega}{\mathrm{c}} = \frac{\omega}{\mathrm{u}} (= \beta)$$

Thus the *wave number* is *2π times the reciprocal of the wave length*. It should also be noted that this wave number is 2π times the one that is used in optics (which is the reciprocal of the wavelength), i.e.

$$k \text{ or } \beta = \frac{\omega}{\mathrm{c}} = \frac{2\pi}{\lambda}$$

Thus it is seen that the *uniform plane wave* has no component in the direction of propagation and is transverse in nature, i.e. the wave has components (of E and H) only in the directions perpendicular to the direction of propagation.

The properties of the polarized light were found to be associated with the directions of **E** and **B** (or **H**); the wave described here is 'plane polarized', the optical 'plane of polarization' being a plane defined by the line of polarization Oz and the direction of **B** or **H**. In referring to the radio waves, however, **E** and not **B** (or **H**) is used to define the direction of polarization, and (since a direction is defined by a line), the preferred description is *linearly polarized*. The cross members in a television aerial indicate the direction of polarization of the transmitted waves.

In some books, a plane wave is defined as a wave in which only the phase, not necessarily the magnitude, of **E** or **B** is constant over a plane perpendicular to the direction of propagation.

12.7 BOUNDARY CONDITIONS (GENERALIZATIONS)

Earlier we have discussed separately the two types of boundary conditions for the electric fields (Chapter 2, Section 2.9), and the boundary conditions for the electric currents (Chapter 6, Section 6.9), and also the conditions for the magnetic fields (Chapter 8, Section 8.7). Since, for the general time-varying electromagnetic fields, both the electric and the magnetic fields coexist, we shall now generalize the boundary conditions for electromagnetic fields. However it should be clearly understood that the two conditions of the electrostatic fields and the two conditions of the magnetostatic fields do **not** produce four independent conditions for the electromagnetic fields. Essentially there are only two independent boundary conditions, i.e. one based on the Gauss' theorem, and the other one based on the line-integral theorem, and the remaining two conditions are always reducible to the former two. A mathematical proof of this statement is given in Appendix 8. We shall now state below, in completely general terms, the definition and the derivation of the boundary conditions.

When there exists discontinuity in the material, the components of the field vectors on the two sides are related by conditions called the 'boundary conditions'. In some cases, a vector may have a zero value on one side, e.g. in an air-gap between the surfaces of iron for which $\mu \to \infty$, $\mathbf{H} = 0$ in iron, so that its surface is the boundary for the field. In theory, we may have infinity as one of the boundaries, and have to consider what happens there.

The surfaces of discontinuity (denoted by Σ) are invariably investigated by two tools.

1. Application of the appropriate flux theorem (Figure 12.8) to a coin-shaped surface S, enclosing a part of Σ (coin-shaped means 'very thin' compared with its lateral dimension). The flux equations are those underlying the equations div $\mathbf{D} = \rho_C$, and div $\mathbf{B} = 0$ (in integral forms), which give information about **D** and **B**.
2. Application of the appropriate line-integral theorem to a suitably directed circuit C, enclosing a part of Σ (Figure 12.8). This means the width is << the length. From the curl equations, i.e.

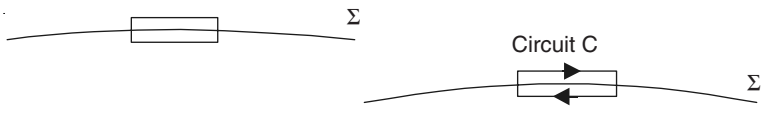


Figure 12.8 A discontinuous interface.

$$\text{curl } \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \qquad \text{and} \qquad \text{curl } \mathbf{H} = \left(\mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}\right) \text{ (in integral forms)}$$

we get information about **E** and **H**.

12.7.1 Some Examples of Generalized Boundary Conditions

Static charge density σ_S on the surface of a perfect conductor

In this condition, the surface Σ separates the region (1) (free space) from the region (2) (conducting) as shown in Figure 12.9. In a perfect conductor, $\mathbf{E} = 0$. This is assumed to imply $\mathbf{D} = 0$. Thus we have no electric field in the region (2), but possibly the components shown in the region (1).

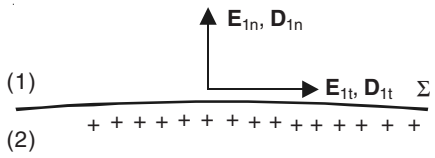


Figure 12.9 Static charge density on the surface of a perfect conductor.

Gauss applied to the coin-shaped surface S on the interface, the cross-sectional area of S being δS, gives

$$D_{1n}\delta S + \text{edge contribution from } D_{1t} = \sigma_S \delta S$$

σ_S being the surface charge density.

But the edge contribution $\to 0$ if the coin is made very thin compared with its width, so we get the boundary condition: $D_{1n} = \sigma_S$.

The electromagnetic induction equation applied to the circuit C (on Σ) supposedly of length δl gives

$$E_{1t}\delta l + \text{edge contribution from } E_{1n} = 0$$

Again the edge contribution $\to 0$, and we get

$$E_{1t} = 0 \tag{12.49}$$

Thus the field is entirely normal and has the value given by

$$D_{1n} = \sigma_S \tag{12.50}$$

Discontinuity of dielectrics (Figure 12.10)

This problem has been discussed for the static fields in Section 2.9 but now here we are generalizing it for the electromagnetic fields. Hence there is no need to assume static conditions.

Gauss applied to the coin-shaped surface S gives

$$(D_{1n} - D_{2n})\delta S + \text{edge contribution from } D_{1t}, D_{2t} = 0,$$

since $\exists$ no charge; therefore, since the edge contribution also $\to 0$

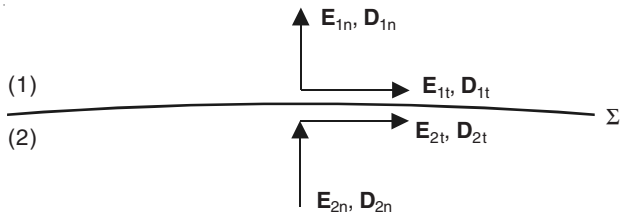


Figure 12.10 Dielectric discontinuity.

$$D_{1n} = D_{2n} \tag{12.51}$$

The electromagnetic induction in a contour of length δl and width $\delta l'$, gives

$$(E_{1t} - E_{2t})\delta l + \text{edge contribution} \propto \delta l' = \left(\frac{\partial B_t}{\partial t}\right)\delta l \delta l'$$

With $\delta l' << \delta l$, only the first term is significant in the above equation, and hence

$$E_{1t} = E_{2t} \tag{12.52}$$

Let the lines of force make angles α_1 and α_2 with the normal to Σ, where

$$\tan \alpha_1 = \frac{E_{1t}}{E_{1n}} \quad \text{and} \quad \tan \alpha_2 = \frac{E_{2t}}{E_{2n}}$$

$$\therefore \quad \frac{\tan \alpha_1}{\tan \alpha_2} = \frac{E_{2n}}{E_{1n}} = \left(\frac{D_{2n}}{\varepsilon_2}\right)\left(\frac{\varepsilon_1}{D_{1n}}\right) \quad \text{and} \quad D_{1n} = D_{2n}$$

$$\therefore \quad \frac{\tan \alpha_1}{\tan \alpha_2} = \frac{\varepsilon_1}{\varepsilon_2} \tag{12.53}$$

Tangential H *on a conducting surface* (Figure 12.11)

We consider the surface of a block of metal of finite conductivity σ. (We neglect the displacement current effects for the present, though our final results will not be affected by their presence.) When an alternating current distribution parallel to the surface is generated in the conducting block (by whatever means, i.e. by induction or from an external source), the current flow is not

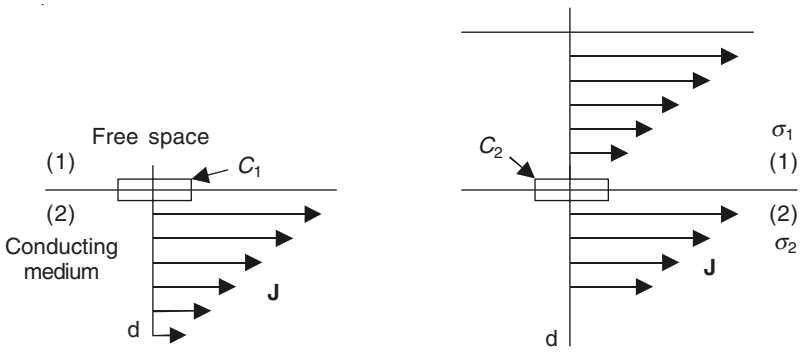


Figure 12.11 Alternating current distributions in conducting media, and interface conditions.

uniformly distributed over the depth, but there is a tendency for the currents to concentrate near the surface due to the skin effect phenomenon, which we shall study in a rigorous manner in Chapter 15. We use the line-integral theorem

$$\oint_C \mathbf{H} \cdot d\mathbf{l} = \iint_S \mathbf{J}_S \cdot d\mathbf{S}$$

i.e. $$\mathbf{n} \times (\mathbf{H}_{t1} - \mathbf{H}_{t2}) = \mathbf{J}_S \leftarrow \text{the surface current density}$$

or $$= 0 \qquad (12.54)$$

if (1) either there is no surface current, or (2) if the currents are distributed in depth in the direction normal to the outer surface or the interface (Figure 12.11). This is because, in the limit, the current enclosed by the contour C_1 or C_2 is infinitesimally small compared with the total current.

PROBLEMS

12.1 State Maxwell's equations for free space, and prove that they are satisfied by

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t}, \quad \mathbf{B} = \text{curl } \mathbf{A}$$

provided that div $\mathbf{A} = 0$, $\quad \nabla^2 \mathbf{A} = \dfrac{1}{c^2} \dfrac{\partial^2 \mathbf{A}}{\partial t^2}$.

Derive $\mathbf{E}$ and $\mathbf{H}$ when

$$\mathbf{A} = \mathbf{i}_x a \cos 2\pi k\,(z - ct) + \mathbf{i}_y b \sin 2\pi k\,(z - ct) + \mathbf{i}_z = 0$$

Verify that $\mathbf{E}$ and $\mathbf{H}$ are orthogonal, and their directions rotate about the z-axis with the frequency kc.

12.2 From the equations:

$$\text{curl } \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}$$

$$\text{curl } \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$

$$\text{div } \mathbf{J} = -\frac{\partial \rho}{\partial t}$$

prove that for time-varying fields div $\mathbf{B} = 0$, and div $\mathbf{D} = \rho$

12.3 Prove that the two divergence Maxwell's equations can be deduced from the two curl equations of Maxwell and the continuity equation.

12.4 Show that the flux of the sum $\left(\mathbf{J} + \dfrac{\partial \mathbf{D}}{\partial t}\right)$ through any closed surface is zero.

12.5 By taking the divergence of one of the Maxwell's equations, show that:

$$\nabla \cdot \mathbf{J} + \frac{\partial \rho}{\partial t} = 0$$

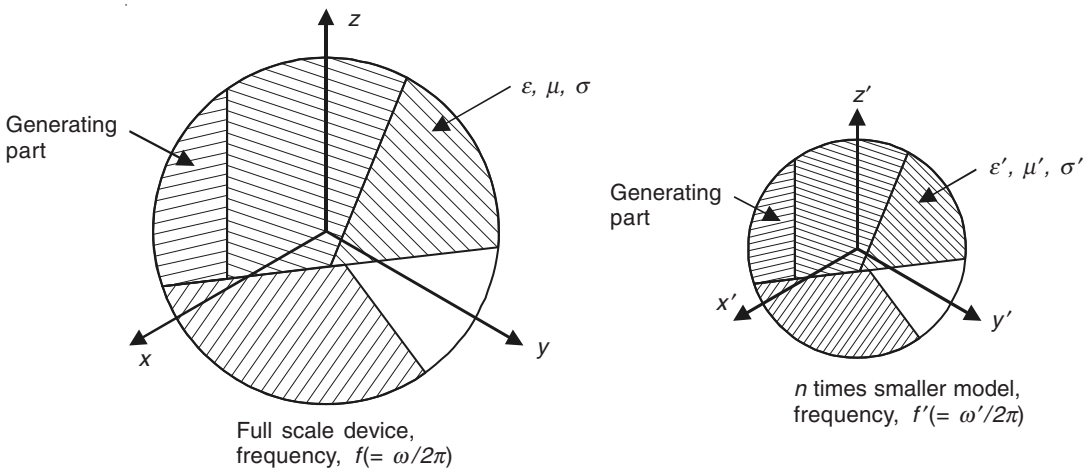
12.6 In a house, where the local AM transmitter is providing a field strength of

$$\mathbf{E} = \mathbf{i}_x \cdot 0.02 \sin [0.1927 (3 \times 10^8 t - z)] \text{ V/m}$$

calculate the rms displacement current density.

Ans.: 0.723 μA/m^2

12.7 Scaled modelling is a normal engineering practice. But since all parameters (physical as well as geometrical) cannot be scaled (up or down) to the same numerical value in a required model, it is necessary to use some figures of merit. Hence for an electromagnetic model, starting from Maxwell's equations, derive the necessary conditions for an electromagnetic field in a small-scale model which has to be similar to the field in a real, n times larger device. (These conditions are often referred to as the 'conditions of electrodynamic similitude'.)



13 Vector Potentials

13.1 INTRODUCTION

At this stage, we shall introduce a 'new' magnetic field vector, which is extremely useful for the proper discussion of inductance problems and also is a highly versatile tool for many other problems. Though we say that this is a new vector, which is denominated as the 'Magnetic vector potential' and its notation being **A**, it has been mentioned briefly in passing [Ref.: Sections 8.2–8.2.2, 10.2–10.2.1, and 11.1.2), we have not established it on a rigorous basis, which we shall do now. One possible way of looking at this vector **A** is that we start from an electromagnetic field (**E**, **B**) in space. This can be explored by means of a test particle of mass m, and charge e moving with a velocity v relative to the observer. The charge e is assumed to be small enough not to disturb the field under study. The test particle possesses a mechanical momentum $m\mathbf{v}$ on account of its mass. But it also possesses electromagnetic momentum, given by $e\mathbf{A}$ on account of its charge, so that the total momentum of the particle is

$$\mathbf{p} = m\mathbf{v} + e\mathbf{A} \tag{13.1}$$

The electromagnetic momentum per unit charge of a test charge is known as the *magnetic vector potential* **A** of the electromagnetic field at the point. We also know that the electric potential energy per unit charge of the test charge at the point is known as the *electric scalar potential V* (or denoted as ϕ sometimes) of the electromagnetic field at the point. This is the physical basis for this new field vector **A**. In fact, it should be noted that it is quite usual to introduce **A** more as a mathematical device and convenience, because it then (apparently?) lacks the physical basis of the potential energy distribution on which the electrostatic scalar potential V or ϕ is based. In that sense, it might be argued that the name 'vector potential' is somewhat of a misnomer. Yet as we shall see that there may be a strong physical basis for this vector, as Maxwell himself expressed his equation of the law of electromagnetic induction in terms of **A** and *not* in terms of **E**. Maxwell called **A** the *electrokinetic momentum vector*. From these considerations, the initial introduction of **A** as an electromagnetic momentum has strong justification.

However we shall now introduce this vector as an adjunct to the set of Maxwell's equations and later show that there are other analogous 'vector potentials' quite useful for solving a wide range of problems.

13.2 MAGNETIC VECTOR POTENTIAL

We start with the Maxwell's electromagnetic field equations, which written in the differential form are

$$\text{div } \mathbf{D} = \nabla \cdot \mathbf{D} = \rho_C \tag{13.2a}$$

$$\text{div } \mathbf{B} = \nabla \cdot \mathbf{B} = 0 \tag{13.2b}$$

$$\text{curl } \mathbf{E} = \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{13.2c}$$

$$\text{curl } \mathbf{H} = \nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \tag{13.2d}$$

Equation (13.2b) is automatically satisfied, if

$$\mathbf{B} = \text{curl } \mathbf{A} = \nabla \times \mathbf{A} \tag{13.3}$$

and it can be shown that any vector whose divergence is zero can be written as the curl of another vector [Ref.: Section 0.7.7]. Thus for any field **B**, **A** can be found (or at least exists).

A is called the vector potential of the magnetic field vector **B**. It is not completely defined (i.e. not uniquely defined) by Eq. (13.3), since if **A′** is any vector of appropriate dimensions whose curl is zero, then,

$$\text{curl } (\mathbf{A} + \mathbf{A}') = \text{curl } \mathbf{A} + \text{curl } \mathbf{A}' = \text{curl } \mathbf{A}$$

so that (**A** + **A′**) will do as well as **A** as a vector potential. We can, in fact, make **A** satisfy another condition such as

$$\text{div } \mathbf{A} = 0 \tag{13.4}$$

[Compare the arbitrary constant in the scalar potential of the electric or the magnetic field, as it (the constant) disappears on differentiation, making no difference to the calculated value of the field.]

Let us now consider a closed circuit or a contour *C* in a magnetic field **B** (Figure 13.1). The flux Φ traversing *C* is given by

$$\Phi = \iint_S \mathbf{B} \cdot d\mathbf{S} = \iint_S (\text{curl } \mathbf{A}) \cdot d\mathbf{S}$$

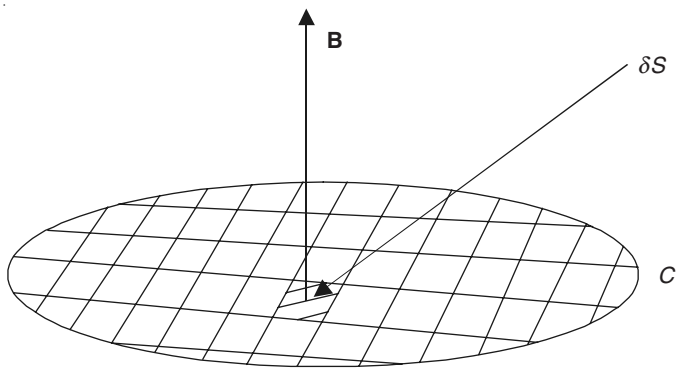


Figure 13.1 A closed contour in a magnetic field **B**.

where S is the area enclosed by C, and δS is one element in a meshwork spanning C. But, by the definition of curl,

$$\text{curl } \mathbf{A} \cdot \delta\mathbf{S} = \text{Line integral of } \mathbf{A} \text{ around the edge of } \delta\mathbf{S}$$

Hence,

$$\begin{aligned}\iint \text{curl } \mathbf{A} \cdot \delta\mathbf{S} &= \text{Sum of line-integrals of } \mathbf{A} \text{ around the edge of all elements } \delta\mathbf{S} \\ &= \text{Line-integral of } \mathbf{A} \text{ round } C. \\ &= \oint_C \mathbf{A} \cdot d\mathbf{l} \qquad \text{(Stokes' theorem)}\end{aligned}$$

i.e. the flux through any closed circuit C is equal to the line integral of $\mathbf{A}$ round C. (13.5)

This shows that the unit of $\mathbf{A}$ is webers/metre.

Let us now consider a magnetic field $\mathbf{B}$ which is fixed in direction at each point, but varies in magnitude with time (Figure 13.2). The associated electric field $\mathbf{E}$ is then related by

$$\text{curl } \mathbf{E} = \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$

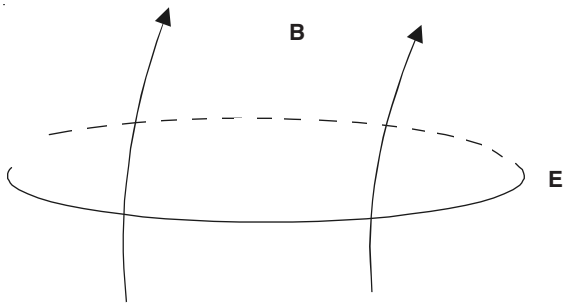


Figure 13.2 An electromagnetic field.

Now, if we write

$$\mathbf{A} = -\int \mathbf{E} \cdot dt \text{ at each point} \tag{13.6}$$

then we shall have

$$\frac{\partial \mathbf{A}}{\partial t} = -\mathbf{E}$$

and

$$\text{curl } \mathbf{E} = -\text{curl}\left(\frac{\partial \mathbf{A}}{\partial t}\right) = -\left(\frac{\partial}{\partial t}\right) \text{curl } \mathbf{A}$$

so that the $\mathbf{A}$ will serve as the vector potential. This gives some hint of a physical interpretation of $\mathbf{A}$. In the particular case, where the magnetic field is varying in magnitude only (e.g. an alternating field), but fixed in direction, $\mathbf{A}$ has the same direction as $\mathbf{E}$.

Earlier we have derived the magnetic fields from a scalar potential Ω, through an equation whose vector form is

$$\mathbf{H} = -\text{grad } \Omega = -\nabla\Omega$$

But curl grad $\Omega = 0$, $\leftarrow$ a vector identity.

Hence, in the magnetic field so defined, we have curl **H** = 0. In general, however,

$$\text{curl } \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}$$

Thus curl **H** ≠ 0 in any region containing either conduction or displacement current. This requirement severely restricts the use of the magnetic scalar potential Ω to determine the magnetic fields. If we write **B** = curl **A** = ∇ × **A**, then since div curl **A** = 0 (← a vector identity), hence ∇ · **B** = 0, which however is a universal property of the magnetic fields; and so it follows that, whereas Ω can be used in special cases only, the vector potential **A** can always be used for evaluation of the magnetic fields. Its special spheres are in problems where Ω is useless, viz:

1. Regions containing conduction currents (**J** ≠ 0);
2. Regions containing displacement currents ← this covers all radiation problems [(∂**D**/∂t) ≠ 0].

General comments. In this section, we have introduced the vector **A** from two different angles, i.e. (a) a vector whose line-integral around a closed contour is equal to the flux linked with the contour; and/or (b) the electric field vector integrated with respect to time gives the vector **A**. Since by this second approach, the time-derivative operator gets eliminated, then **A** exists not only for the constant fields but also for all fields. The second approach further gives a physical basis for **A** whereas the line-integral approach does not make obvious any physical basis for **A**. However it should also be noted that it is possible to use either of the two approaches as a starting point, and derive the other from it.

13.3 CALCULATION OF MAGNETIC VECTOR POTENTIAL IN SOME IMPORTANT CASES

13.3.1 Vector Potential of the Field of Current in a Circuit

We know from Biot–Savart that the magnetic field of an arbitrary circuit can be made from the contributions,

$$\delta H = \frac{i \delta l \sin \theta}{4\pi r^2} \tag{13.7}$$

where $\delta\mathbf{H}$ is perpendicular to both $\delta\mathbf{l}$ and **r** [Figure 13.3(a)].

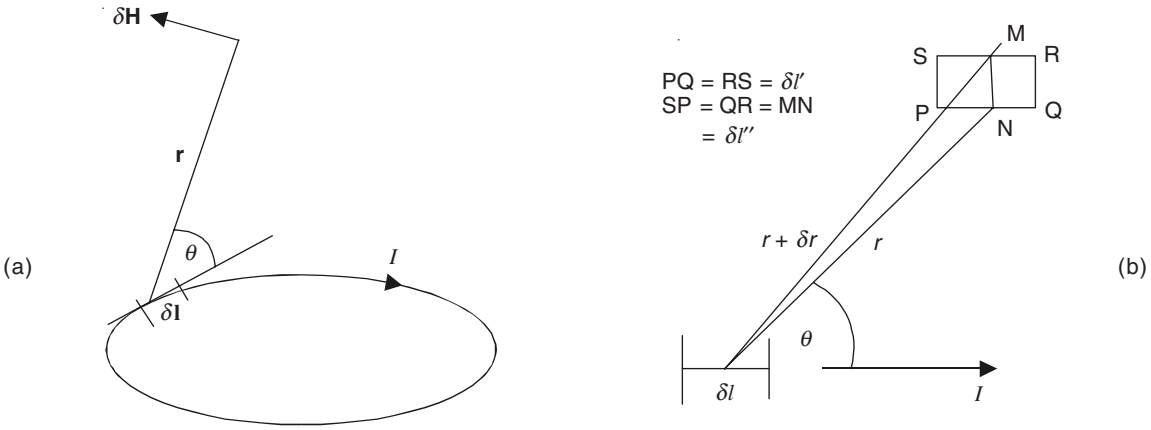


Figure 13.3 A current circuit and its element δl.

We shall now show that this contribution can be deduced by supposing the element δl to set up a vector potential whose direction at every point is the direction of δl, and whose magnitude is

$$\left(\frac{\mu_0}{4\pi}\right)\left(\frac{i\delta l}{r}\right) \tag{13.8}$$

Let us consider a small rectangular circuit PQRS, shown in the Figure 13.3(b), in the plane of δl and r. On PQ, **A** has the mean value (assuming the above expression) of

$$\left(\frac{\mu_0}{4\pi}\right)\left(\frac{i\delta l}{r}\right)$$

and on RS of

$$\left(\frac{\mu_0}{4\pi}\right)\left(\frac{i\delta l}{r+\delta r}\right)$$

To calculate the flux through PQRS, we form

$$\oint_{\text{PQRS}} A_1 dl$$

$\exists$ no contribution from QR or SP, so:

$$\oint_{\text{PQRS}} A_1 dl = \left(\frac{\mu_0}{4\pi}\right) i\delta l\left[\frac{1}{r} - \frac{1}{r+\delta r}\right]\delta l'$$

$$= \frac{\left(\frac{\mu_0}{4\pi}\right)(i\delta l)(\delta r \delta l')}{r^2}, \text{ when } \delta r \text{ is small.}$$

But MN = $\delta l'' = \dfrac{\delta r}{\sin\theta}$; hence

$$\oint_{\text{PQRS}} A_1 dl = \left[\frac{\left(\frac{\mu_0}{4\pi}\right) i\delta l \sin\theta}{r^2}\right]\delta l' \delta l'' \tag{13.9}$$

Thus the bracketed quantity in the above Eq. (13.9) is same as the flux density deduced from Biot–Savart. It may be verified that the assumed **A** gives no component of **B** in any other direction. Thus the assumed formula for **A** which may be expressed by the vector equation

$$\delta\mathbf{A} = \left(\frac{\mu_0}{4\pi}\right)\left(\frac{i\delta\mathbf{l}}{r}\right) \tag{13.10}$$

is consistent with Biot–Savart.

Extrapolating this expression to the surface current, we get

$$\delta\mathbf{A} = \left(\frac{\mu_0}{4\pi}\right)\left(\frac{J_S \delta\mathbf{S}}{r}\right) \tag{13.11}$$

and to the volume current

$$\delta\mathbf{A} = \left(\frac{\mu_0}{4\pi}\right)\left(\frac{J_v \delta v}{r}\right) \tag{13.12}$$

The theory of this section applies to the steady-state conditions. In the next few sections, it is going to be applied to the transmission lines, with particular reference to the travelling waves—which is a quasi-steady or non-steady problem. Then we shall be answering the question regarding the justification of this extension.

Later it is also shown (in Sections 19.2.1 to 19.2.2, on oscillating dipoles) that

$$\delta \mathbf{A} = \left(\frac{\mu_0}{4\pi}\right)\left(\frac{[i]\,\delta \mathbf{l}}{r}\right) \tag{13.13}$$

where $[i]$ is a retarded value of the current. Thus **A** is related to the conduction current only, provided the retarded values are taken (see Section 13.6 for the explanation of the retardation effects).

The same argument applies to the relation of the scalar potential V with the charges. Thus although the parallel between **A** and V, in the subsequent sections, is developed for the quasi-steady cases, it does in fact hold good in dynamic cases. For example, the conclusion ($LC = 1/c^2$) is generally valid.

13.3.2 Systems of Parallel Current-carrying Conductors

If the currents are all parallel to one direction, **A** has only one component, and is virtually a scalar.

13.3.2.1 A single short conductor

As shown in Figure 13.4, a straight conductor AB of finite length $(b - a)$, is carrying a current i.

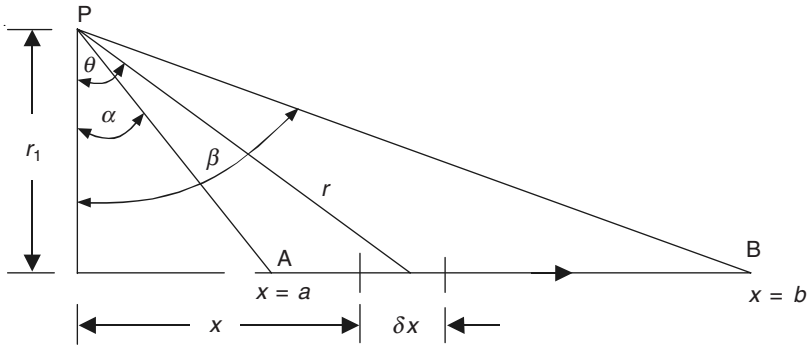


Figure 13.4 A short straight conductor.

We wish to find the vector potential at a point P, distant r_1 from the line of the conductor. Hence

$$A_P = \left(\frac{\mu_0 i}{4\pi}\right)\int_{x=a}^{x=b} \frac{dx}{r}$$

We change the variable to θ, so that $x = r_1 \tan\theta$, $dx = r_1 \sec^2\theta d\theta$, $r = r_1 \sec\theta$, and the limits of θ are α and β corresponding to $x = a$ and $x = b$, respectively, and so

$$A_P = \left(\frac{\mu_0 i}{4\pi}\right)\int_{\alpha}^{\beta} \sec\theta\, d\theta = \left(\frac{\mu_0 i}{4\pi}\right)\left[\ln(\sec\theta + \tan\theta)\right]_{\alpha}^{\beta}$$

Also,

$$\sec\alpha = \frac{\sqrt{r_1^2 + a^2}}{r_1}, \quad \tan\alpha = \frac{a}{r_1}$$

and

$$\sec\beta = \frac{\sqrt{r_1^2 + b^2}}{r_1}, \quad \tan\beta = \frac{b}{r_1}$$

$$\therefore \quad \mathbf{A}_P = \mathbf{i}_x \left(\frac{\mu_0 i}{4\pi}\right) \ln\left(\frac{\sqrt{r_1^2 + b^2} + b}{\sqrt{r_1^2 + a^2} + a}\right) \tag{13.14}$$

13.3.2.2 A pair of long conductors (go and return)

We extrapolate from the last problem, and referring to Figure 13.4, we move the point P (under consideration) to a plane which bisects the conductor AB normally, so that $a = -b$, and then

$$\mathbf{A}_P = \mathbf{i}_x \left(\frac{\mu_0 i}{4\pi}\right) \ln\left(\frac{\sqrt{r_1^2 + b^2} + b}{\sqrt{r_1^2 + b^2} - b}\right)$$

$$= \mathbf{i}_x \left(\frac{\mu_0 i}{2\pi}\right) \ln\left(\frac{\sqrt{r_1^2 + b^2} + b}{r_1}\right) \tag{13.15}$$

At this stage, we cannot let $b \to \infty$, because this would make A also $\to \infty$. So we take a conductor pair, with currents $\pm i$ and the distance of the point P from them being r_1 and r_2 respectively (Figure 13.5). We get

$$\mathbf{A}_P = \mathbf{i}_x \left(\frac{\mu_0 i}{2\pi}\right) \ln\left(\frac{\sqrt{r_1^2 + b^2} + b}{r_1}\right)\left(\frac{r_2}{\sqrt{r_2^2 + b^2} + b}\right)$$

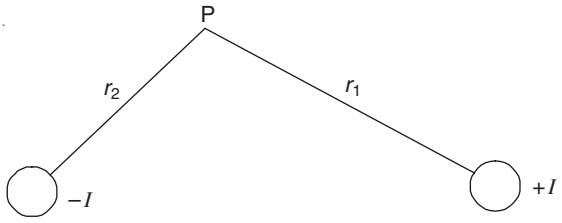


Figure 13.5 A pair of parallel long conductors.

Now, we permit b to become larger and larger, so as $b \to \infty$,

$$\mathbf{A}_P \to \mathbf{i}_x \left(\frac{\mu_0 i}{2\pi}\right) \ln\left(\frac{r_2}{r_1}\right) \tag{13.16}$$

Hence from this we can deduce the potential for the single conductor.

13.3.2.3 A single infinitely long conductor

This gives directly

$$\mathbf{A} = -\mathbf{i}_x \left(\frac{\mu_0 i}{2\pi} \right) \ln r \tag{13.17}$$

at a point distant r from the conductor. As with the scalar potentials, a constant can be added without affecting the validity of the formula. Now, we apply a check by calculating the magnetic flux density. For this, we use the cylindrical polar coordinate system with the current and the conductor in the z-direction. Hence, in the (r, ϕ, z) system, we have

$$\mathbf{A} = \mathbf{i}_r A_r + \mathbf{i}_\phi A_\phi + \mathbf{i}_z A_z = \mathbf{i}_r 0 + \mathbf{i}_\phi 0 + \mathbf{i}_z \left[-\left(\frac{\mu_0 i}{2\pi} \right) \ln r \right]$$

and

$$\text{curl } \mathbf{A} = \mathbf{i}_r \left[\frac{1}{r} \frac{\partial A_z}{\partial \phi} - \frac{\partial A_\phi}{\partial z} \right] + \mathbf{i}_\phi \left[\frac{\partial A r}{\partial z} - \frac{\partial A_z}{\partial r} \right] + \mathbf{i}_z \left[\frac{\partial A_\phi}{\partial r} - \frac{1}{r} \frac{\partial A_r}{\partial \phi} \right]$$

from which

$$\mathbf{B} = \mathbf{i}_r B_r + \mathbf{i}_\phi B_\phi + \mathbf{i}_z B_z = \mathbf{i}_r 0 + \mathbf{i}_\phi \left(\frac{\mu_0 i}{2\pi r} \right) + \mathbf{i}_z 0 \tag{13.18}$$

which we know to be true.

The resemblance between the magnetic vector potential for a line current and the formula for the scalar potential of a charged wire, which is

$$V = -\left(\frac{Q}{2\pi\varepsilon_0} \right) \ln r \tag{13.19}$$

is striking and not accidental. Thus, if we take two wires of radius a, spacing b, and carrying the charges $\pm Q$ coulombs per unit length (Figure 13.6), then the electrostatic potential at P is

$$V_{\mathrm{P}} = \left(\frac{Q}{2\pi\varepsilon_0} \right) \left[\ln \left(\frac{r'}{r} \right) \right] \tag{13.20}$$

Moving the point P to the surface of the charged conductors successively, we get

$$\text{on the +ve wire, } V_+ = \left(\frac{Q}{2\pi\varepsilon_0} \right) \left[\ln \left(\frac{b}{a} \right) \right]$$

$$\text{on the } -\text{ve wire, } V_- = \left(\frac{Q}{2\pi\varepsilon_0} \right) \left[\ln \left(\frac{a}{b} \right) \right]$$

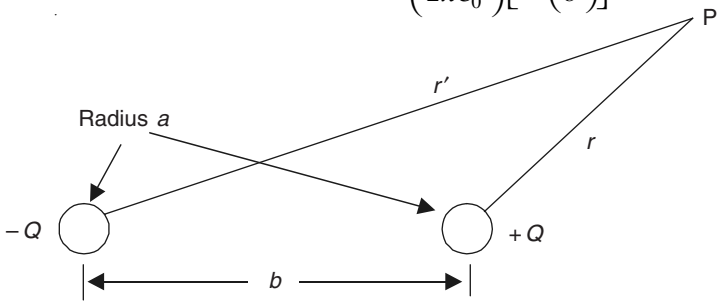


Figure 13.6 Two parallel line charges.

$$\therefore \quad \text{The capacitance } C \text{ per unit length} = \frac{Q}{V_+ - V_-} = \frac{\pi\varepsilon_0}{\ln\left(\frac{b}{a}\right)} \tag{13.21}$$

Now let the same wires carry currents $\pm i$. Similar to the case of the line charges, we have at P the magnetic vector potential as

$$A = \frac{\mu_0 i}{2\pi}\left[\ln\left(\frac{r'}{r}\right)\right] \tag{13.22}$$

Hence on the conductor surfaces, A will be

$$A = \pm\frac{\mu_0 i}{2\pi}\left[\ln\left(\frac{b}{a}\right)\right] \tag{13.23}$$

but we have seen from Eq. (13.5),

the change in **A** between one conductor and the other = the flux between them.

$$\text{Hence, } \Phi = \left(\frac{\mu_0 i}{\pi}\right)\left[\ln\left(\frac{b}{a}\right)\right] \text{ per unit length} \tag{13.24}$$

and

$$\therefore \quad \text{the inductance } L \text{ per unit length} = \left(\frac{\mu_0}{2\pi}\right)\left[\ln\left(\frac{b}{a}\right)\right] \tag{13.25}$$

(*Note:* The above value of self-inductance neglects the internal self-inductance of the wires.)

$\therefore$ From Eqs. (13.21) and (13.25),

$$LC = \mu_0\varepsilon_0 = \frac{1}{c^2} \tag{13.26}$$

It can be shown that the travelling wave on the line, travels with the velocity $(LC)^{-1/2}$. Equation (13.26) shows that, for an air-insulated line, this is the velocity of light.

It is to be further noted that if the two points lie on the same line of force, there is no flux between them, and so there is no change in **A**, i.e. in a two-dimensional field,

$$\mathbf{A} = \text{constant, is a line of force} \tag{13.27}$$

The above applies to more complex systems of parallel conductors, with extensive analogues with the electric case. For example, charged wires over a perfectly conducting earth (Figure 13.7) give rise to a field ascribed to image charges which make the earth equipotential. Time-varying currents give similar images; for if the flux penetrated the earth, it would give rise to eddy

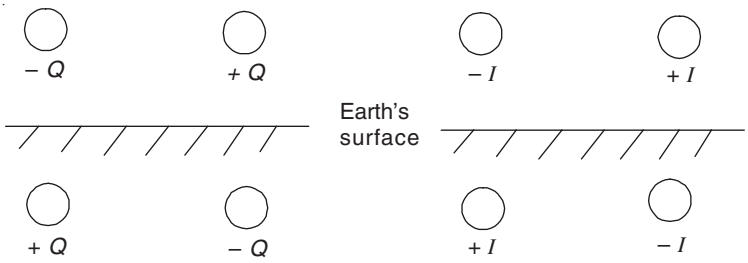


Figure 13.7 Charges and line currents above the equipotential earth surface, and their equivalent images.

currents, tending to exclude it (the flux). Thus the earth's surface, if perfectly conducting, is a line of force. Hence, we can get L for the conductors in presence of the earth.

(The theory of eddy currents—due to finite conductivity—and their effects will be studied subsequently in Chapter 15.)

13.4 INDUCTANCE IN TERMS OF VECTOR POTENTIAL

We have defined inductance earlier both in terms of the flux-linkage (Sections 10.6 to 10.6.3) as well as in terms of the stored energy (Sections 11.15 to 11.15.1), and also considered simple configurations in terms of the vector potential in the previous two sections. Now we shall derive the general expressions for inductance in terms of the magnetic vector potential.

The self-inductance of a circuit is given by

$$L = \left(\frac{1}{I^2}\right)\sum \Phi \delta i \qquad \text{[see Eq. (11.64)]}$$

where δi is a filament of current, and Φ is the flux linked with that current filament.

Let $\mathbf{A}$ be the magnetic vector potential set up by the whole current I. Then

$$\Phi = \oint_C \mathbf{A} \cdot d\mathbf{l} \qquad \text{[see Eq. (13.5)]}$$

where C is the contour of the particular filament for which the flux is being calculated (Figure 13.8). Also, if $\mathbf{J}$ is the current density in the filament, and $\delta\mathbf{S}$ is the cross-sectional area of the filament, then

$$\delta i = \mathbf{J} \cdot \delta\mathbf{S}$$

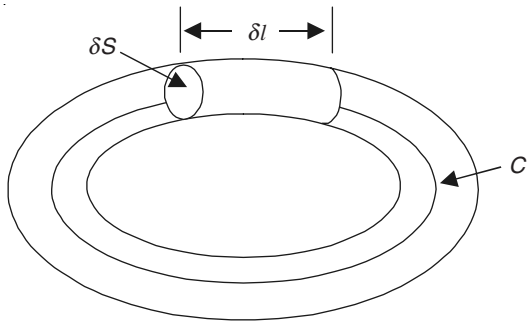


Figure 13.8 A current filament.

∴ The relation for the inductance can be written as

$$L = \left(\frac{1}{I^2}\right)\sum \left[\oint_C A_1 dl\right] \mathbf{J} \cdot d\mathbf{S} \tag{13.28}$$

In this expression, the flux-linkage with a particular filament C is obtained by adding up ($\Sigma A_1 \delta l$) for each element δl along the length of the filament, and then it is multiplied by $\mathbf{J} \cdot d\mathbf{S}$, and then a second summation is performed to take account of every filament in the wire. In the associated figure (Figure 13.8), the cross-section of the wire has been greatly exaggerated for the purpose of

clarity. In reality, the diameter of the wire is quite small compared with the length dimension of the circuit. So the expression for the inductance [i.e. Eq. (13.28)] can be further simplified. This is done by assuming the wire to be divided into short slices in which δl is same for all filaments, and reverse the order of summation; i.e. first add for all the filaments in the slice, and then for all the slices in the circuit. Hence,

$$L = \left(\frac{1}{I^2}\right)\sum\left[(\Sigma\, A_1 J\, dS)\, dl\right] = \Sigma(L' dl) \tag{13.29}$$

where

$$L' = \left(\frac{1}{I^2}\right)\sum(A_1 J\, dS) = \left(\frac{1}{I^2}\right)\iint_S (A_1 J\, dS) \tag{13.30}$$

This integration is over all the elements of the area in the cross-section. If the current distribution is uniform over the whole cross-sectional area, then, $J = I/S$, and

$$L' = \left(\frac{1}{IS}\right)\iint_S (A_1 dS) \tag{13.31}$$

The significance of this derivation is that the inductance L which has been considered so far as a property of the circuit as a whole (as a consequence of the total flux-linkage) is now associated with $(L'dl)$. L' is now defined in terms of **A**, and hence is not unique as **A** itself is not unique. But $\Sigma(L'dl)$ added all around the circuit, results in a unique sum. Also, since the vector potential due to a current in a conductor of any section, is same as if the current were concentrated in a filament, we can use Eqs. (13.31) and (13.29) to calculate the inductance L.

13.4.1 Mutual Inductance

Similar expressions as in the previous section can be derived for the mutual inductance. Now there are two circuits (1) and (2). So we find the vector potential $\mathbf{A}_2$ set up by the current I_2 in the circuit (2), and the mutual inductance between the two circuits M_{12}, which is given as

$$M_{12} = \Sigma\,(M'_{12} dl_1) \tag{13.32}$$

where

$$M'_{12} = \left(\frac{1}{I_1 I_2}\right)\iint_{S_1} A_{21} J_1 dS_1 \tag{13.33}$$

I_1 and I_2 being the currents in the circuits (1) and (2) respectively. If the currents are uniformly distributed, then

$$M'_{12} = \left(\frac{1}{I_2 S_1}\right)\iint_{S_1} A_{21} dS_1 \tag{13.34}$$

If the spacing between the two circuits is large enough to allow $\mathbf{A}_2$ to be regarded constant over the cross-section of the conductor in (1), then we get

$$M'_{12} = \frac{A_{21}}{I_2} \tag{13.35}$$

Note that $\mathbf{A}_{21}$ is the component of the vector potential of I_2 which is tangential to the element dl_1 of the circuit (1).

13.4.2 Neumann's Formula for Mutual Inductance

This formula was derived by Franz Neumann, a German mathematician, in 1845, through the concept of the flux-linkage and the vector potential.

In Figure 13.9, a current I_1 in the circuit (1) produces a flux Φ_{12} which links the circuit (2). The mutual inductance M_{12} is

$$M_{12} = \frac{\Phi_{12}}{I_1} \tag{13.36}$$

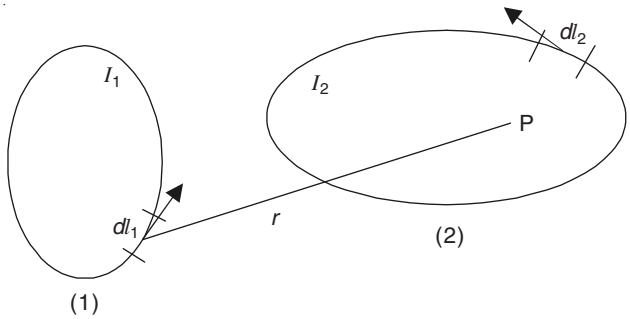


Figure 13.9 Mutual inductance between two loops.

The vector potential at the point P due to the current I_1 in the circuit (1) is

$$\mathbf{A} = \left(\frac{\mu_0 I_1}{4\pi}\right) \oint_1 \frac{d\mathbf{l}_1}{r}$$

The corresponding flux density at P is

$$\mathbf{B} = \text{curl } \mathbf{A} = \nabla \times \mathbf{A}$$

$\therefore$ The total flux through the circuit (2) is

$$\Phi = \iint_2 \mathbf{B} \cdot d\mathbf{S} = \iint_2 (\nabla \times \mathbf{A}) \cdot d\mathbf{S}$$

By Stokes' theorem,

$$\iint_2 (\nabla \times \mathbf{A}) \cdot d\mathbf{S} = \oint_2 \mathbf{A} \cdot d\mathbf{l}_2$$

$$\therefore \quad \Phi_{12} = \oint_2 \mathbf{A} \cdot d\mathbf{l}_2 = \left(\frac{\mu_0 I_1}{4\pi}\right) \oint_1 \oint_2 \frac{d\mathbf{l}_1 \cdot d\mathbf{l}_2}{r}$$

$$\therefore \quad M_{12} = \left(\frac{\mu_0}{4\pi}\right) \oint_1 \oint_2 \frac{d\mathbf{l}_1 \cdot d\mathbf{l}_2}{r} \tag{13.37}$$

Similarly a current I_2 in the circuit (2) produces a flux Φ_{21} which links the circuit (1), and is given by

$$\Phi_{21} = \oint_1 \mathbf{A} \cdot d\mathbf{l}_1 = \left(\frac{\mu_0 I_2}{4\pi}\right) \oint_2 \oint_1 \frac{d\mathbf{l}_2 \cdot d\mathbf{l}_1}{r}$$

Hence the mutual inductance is the same as given by Eq. (13.37), and

$$\therefore \qquad M_{12} = M_{21}$$

13.4.3 Mutual Inductance between Parallel Conductors of Length l

We consider two parallel filaments of the same length l as shown in Figure 13.10. From the dimensions indicated on the figure, we have

$$\frac{d\mathbf{l}_2 \cdot d\mathbf{l}_1}{r} = \frac{dy_1 dy_2}{\left[b^2 + (y_1 - y_2)^2\right]^{1/2}}$$

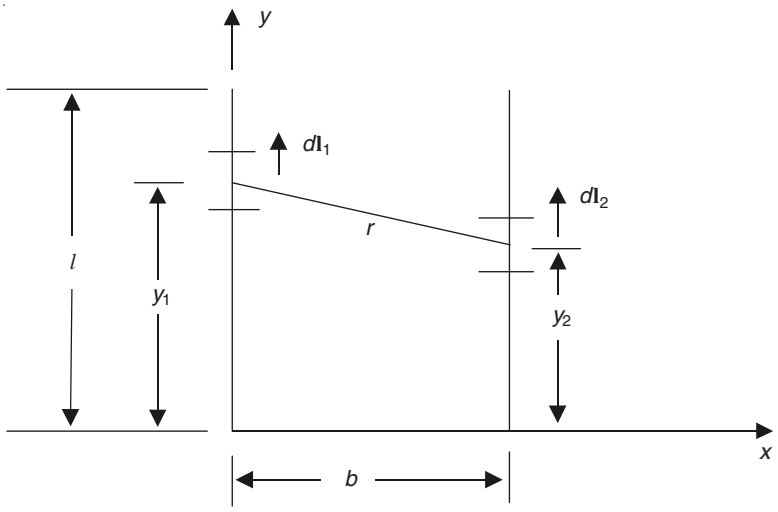


Figure 13.10 Parallel conductors of length l.

Hence the integral of Eq. (13.37) becomes, on using the substitution $(y_1 - y_2) = z$:

$$\int_0^l dy_1 \int_0^l \frac{dy_2}{\left[b^2 + (y_1 - y_2)^2\right]^{1/2}} = \int_0^l dy_1 \int_{y_1}^{y_1 - l} \frac{dz}{\left(b^2 + z^2\right)^{1/2}}$$

$$= \int_0^l -\sinh^{-1}\left(\frac{z}{b}\right)\Big|_{y_1}^{y_1 - l} dy_1 = \int_0^l \left[-\sinh^{-1}\left(\frac{y_1 - l}{b}\right) + \sinh^{-1}\left(\frac{y_1}{b}\right)\right] dy_1$$

$$= b\left[\left(\frac{y_1}{b}\right)\sinh^{-1}\left(\frac{y_1}{b}\right) - \sqrt{1 + \left(\frac{y_1}{b}\right)^2}\right]\Bigg|_0^l$$

$$- b\left[\left(\frac{y_1 - l}{b}\right)\sinh^{-1}\left(\frac{y_1 - l}{b}\right) - \sqrt{1 + \left(\frac{(y_1 - l)}{b}\right)^2}\right]\Bigg|_0^l$$

$$\therefore \qquad M_{12} = \left(\frac{\mu_0}{4\pi}\right) \oint_1 \oint_2 \frac{d\mathbf{l}_1 \cdot d\mathbf{l}_2}{r}$$

$$= \left(\frac{\mu_0 l}{2\pi}\right)\left[\sinh^{-1}\left(\frac{l}{b}\right) - \sqrt{1+\left(\frac{b}{l}\right)^2} + \left(\frac{b}{l}\right)\right] \tag{13.38}$$

This equation can be used to evaluate the mutual inductance between two rectangular coils by using the principle of superposition.

13.4.4 Self-inductance of a Pair of Parallel Conductors of any Cross-section

We consider a go-and-return circuit of a pair of long straight parallel conductors. The sections of the two conductors are similar and also are similarly oriented. Since **A** now has only one component, parallel to the direction of the currents, we can drop the suffix 1 of Eq. (13.31) which we shall use for calculation. We subdivide **A** into two components A_1 and A_2 for the currents in the two conductors (1) and (2) as shown in Figure 13.11. Then, Eq. (13.31) applied to this configuration becomes

$$L' = \frac{1}{IS}\iint (A_1 + A_2)\, dS \tag{13.39}$$

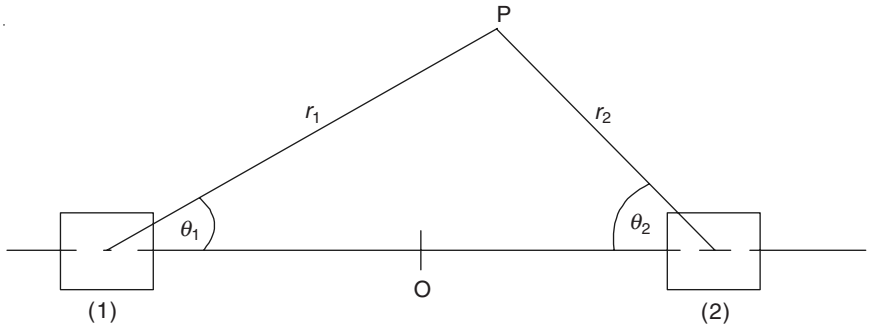


Figure 13.11 Inductance of a pair of parallel long conductors of any cross-section.

Since the circuit is symmetrical, we can choose the elements such that their contributions to the self-inductance are the same for each conductor. Since the currents in the two conductors are equal and opposite, the arbitrary constant in A be so assigned that the values of A at the corresponding points in the two conductors are equal and opposite. Hence A must be zero at O, the centre of symmetry. Hence, if the vector potential at P due to (1) is given by

$$A_1 = f(r_1, \theta_1)$$

then A_2 must be

$$A_2 = -f(r_2, \theta_2) \tag{13.40}$$

so that at O, $r_1 = r_2$ and $\theta_1 = \theta_2$, which would make

$$A_1 + A_2 = 0 \text{ at O.}$$

13.4.5 Rods of Circular Cross-section

We shall consider a pair of parallel straight rods of radius a and spacing b, carrying currents in

opposite directions. For this purpose, we shall derive the vector potential due to a single straight rod of radius a and carrying a current I. The potential outside the rod is same as that due to a line filament, which we found to be [from Eq. (13.17)] as

$$A = C - \left(\frac{\mu_0 I}{2\pi}\right) \ln r, \qquad r > a$$

where A is parallel to the direction of the current. To find A inside the rod, we remind ourselves that the magnetic flux density inside is circumferential, but its magnitude is

$$B = \left(\frac{\mu_0 I r}{2\pi a^2}\right)$$

and remembering from Eq. (13.18) that

$$B = -\frac{\partial A}{\partial r}$$

which on integrating gives us

$$A = C' - \left(\frac{\mu_0 I r}{4\pi a^2}\right)$$

We have to adjust the constants C and C', so that A remains continuous on the surface of the conductor, i.e. $r = a$. If we choose A to be zero on this surface, then

$$\begin{aligned} A_i &= \left(\frac{\mu_0 I}{4\pi a^2}\right)(a^2 - r^2) \qquad r < a \\ A_o &= \left(\frac{\mu_0 I}{2\pi}\right) \ln\left(\frac{r}{a}\right) \qquad r > a \end{aligned} \tag{13.41}$$

The variation of A with r is shown in Figure 13.12. With the boundary condition used, A has the same direction as I inside the conductor, but opposite that of the current outside it.

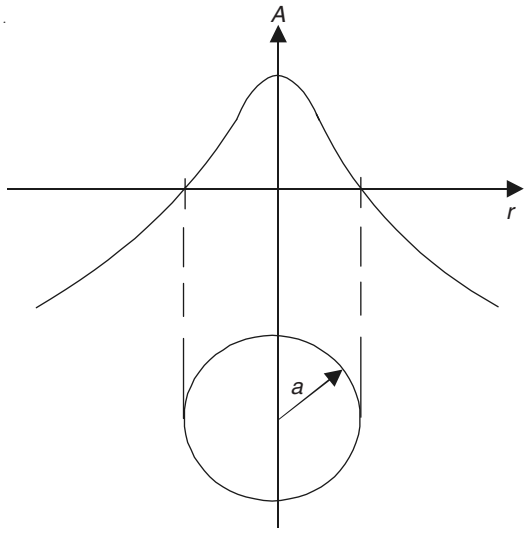


Figure 13.12 Vector potential of a single straight conductor of circular cross-section.

Now, we consider the two parallel conductors, as shown in Figure 13.13. The vector potential components, similar to the problem in Section 13.4.4 are

$$A_1 = \left(\frac{\mu_0 I}{4\pi a^2}\right)(a^2 - r_1^2) \text{ inside the conductor } C_1,$$

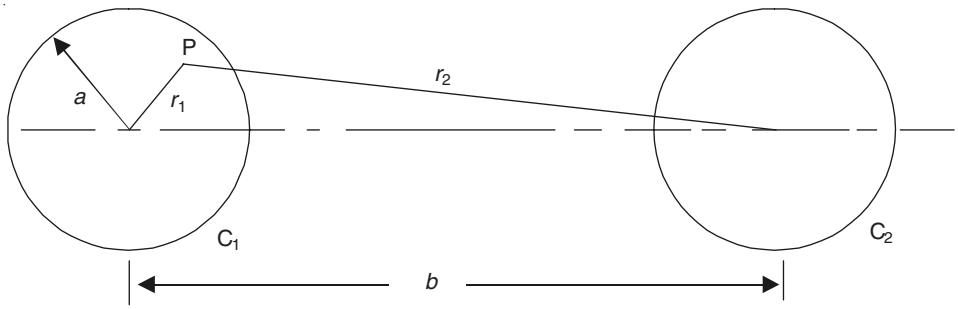


Figure 13.13 Two parallel conductors of circular cross-section, carrying the same current I in opposite directions.

and

$$A_1 = -\left(\frac{\mu_0 I}{2\pi}\right)\ln\left(\frac{r_1}{a}\right) \text{ outside the conductor } C_1,$$

whilst

$$A_2 = +\left(\frac{\mu_0 I}{2\pi}\right)\ln\left(\frac{r_2}{a}\right) \text{ outside the conductor } C_2.$$

As before, $A_1 + A_2 = 0$ at the point O.

Hence, we get L' as

$$L' = \left(\frac{1}{\pi a^2 I}\right)\left[\iint\left(\frac{\mu_0 I}{4\pi a^2}\right)(a^2 - r_1^2)\,dS + \iint\left(\frac{\mu_0 I}{2\pi}\right)\ln\left(\frac{r_2}{a}\right)dS\right]$$

This has to be integrated over the cross-section of the conductor C_1.

By taking a thin circular annulus as the surface element, we get

$$dS = 2\pi r_1 dr_1$$

which gives the first integral as $= \left(\frac{\mu_0 I}{4\pi a^2}\right)\int_0^a (a^2 - r_1^2)\,2\pi r_1 dr_1$

$$= \frac{\mu_0 I a^2}{8}$$

$$\text{and the second integral} = \pi a^2 \times \left(\frac{\mu_0 I}{2\pi}\right)\ln\left(\frac{R_2}{a}\right)$$

where R_2 is the geometric mean distance between the centre of C_2 and the cross-section of C_1. Note that $R_2 \simeq b$.

$\therefore$ Combining both the terms,

$$L' = \left(\frac{\mu_0}{2\pi}\right)\left[\ln\left(\frac{b}{a}\right) + \frac{1}{4}\right] \tag{13.42}$$

This is the self-inductance per metre length of the conductor, and the self-inductance of the circuit will be double this quantity which is what was seen in Section 11.5.1.

13.4.6 The Inductance of a System of Several Circuits

We consider a group of N conductors (circuits) surrounded by air or other non-magnetic materials. The flux-linkage in any circuit is due to the combined effect of the current in that circuit and in all others, so that the flux-linkages can be expressed as

$$\begin{aligned}
\Phi_1 &= L_{11}i_1 + M_{12}i_2 + M_{13}i_3 + \cdots + M_{1N}\mathbf{i}_N \\
\Phi_2 &= M_{21}i_1 + L_{22}i_2 + M_{23}i_3 + \cdots + M_{2N}\mathbf{i}_N \\
&\;\;\vdots \qquad\quad \vdots \qquad\quad \vdots \\
\Phi_N &= M_{1N}i_1 + M_{2N}i_2 + M_{3N}i_3 + \cdots + L_{NN}\mathbf{i}_N
\end{aligned} \tag{13.43}$$

In deriving these N equations, generally we think of N electrically distinct circuits. But, in reality, there are times when the same conductor belongs to more than one circuit. For this purpose, we consider a three-phase transmission line with a neutral conductor, which consists of four parallel conductors (and a dummy return) as shown in Figure 13.14. Such a system really consists of six

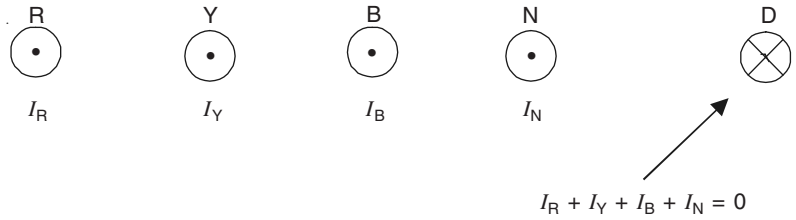


Figure 13.14 Three-phase transmission line with a neutral and a dummy return.

circuits, six being the number of ways of pairs up the conductors. These six circuits would have six self-inductances, and fifteen mutual inductances, which can be associated with the elements of lengths in the conductors (as per Section 13.4). These 21 inductances can then be expressed in terms of four self-inductances of the conductors and six mutual-inductances of the conductor pairs.

Let the currents in the four conductors R, Y, B, N of Figure 13.14 be i_R, i_Y, i_B, i_N, respectively, considered to be positive when coming out of the plane of the paper . Since we have shown no return path, it implies that at every instant,

$$i_R + i_Y + i_B + i_N = 0 \tag{13.44}$$

For the purpose of our calculation, we imagine a 'dummy return conductor D' *in any position*. We shall see that this dummy need not have any physical existence. We imagine that the current i_R circulates in the circuit (R, D); the current i_Y in the circuit (Y, D); the current i_B in the circuit (B, D); and i_N in the circuit (N, D). The total current in the dummy would be $(i_R + i_Y + i_B + i_N)$ which is always zero and hence D does not need to have a physical existence. The self-inductance L_{RR} of the circuit (R, D) is called the 'self-inductance of the conductor R'; and the mutual inductance M_{RY} between the circuits (R, D) and (Y, D), is called the 'mutual inductance between the conductors R and Y'. Thus we have defined L_{RR} and M_{RY} quite unambiguously.

Now we consider the circuit (R, Y). Its self-inductance is set up by a current of +1 amp in R and of −1 amp in Y. This linkage is equivalent to that by a current of +1 amp in (R, D) which is $(L_{RR} - M_{RY})$; and that by a current of −1 amp in (Y, D) which is $(L_{YY} - M_{RY})$. Hence the self-inductance of the circuit (R, Y) is then given by

$$L_{(self)} = L_{RR} + L_{YY} - 2M_{RY} \tag{13.45}$$

Similarly, the flux-linkage set up in (B, N) by a current of +1 amp in (R, D) is $(M_{RB} - M_{RN})$; and that set up by a current of −1 amp in (Y, D) is $(M_{YN} - M_{YB})$. Thus the mutual inductance between the circuits (R, Y) and (B, N) is thus

$$M_{(mutual)} = M_{RB} + M_{YN} - M_{RN} - M_{YB} \tag{13.46}$$

Similarly, the mutual inductance between the circuits (R, Y) and (R, B) can be deduced and be shown to be:

$$M_{(mutual)} = L_{RR} + M_{YB} - M_{RY} - M_{RB} \tag{13.47}$$

This is how the inductance associated with various circuits can be derived from the knowledge of the inductances associated with the conductors. The conductor inductances (both self- and mutual) depend on the location of the dummy conductor, but when they are combined in the equations of the types shown in (13.45) to (13.47), the terms depending on the location of the dummy conductor D will disappear.

Similarly, we can obtain the equivalent equations of a three-wire system, which will be comparatively simpler and easier.

13.5 APPLICATION OF THE MAGNETIC VECTOR POTENTIAL IN TIME-VARYING FIELDS

So far we have considered the magnetic vector potential for the time-invariant magnetic fields, in the rigorous sense. Now we shall generalize the concept and cover its application to the whole range of electromagnetic fields, i.e. the time-varying fields as well. Hence, we start again from the Maxwell's equations which we write in differential form as

$$\text{div } \mathbf{D} = \nabla \cdot \mathbf{D} = \rho_C \tag{13.2a}$$

$$\text{div } \mathbf{B} = \nabla \cdot \mathbf{B} = 0 \tag{13.2b}$$

$$\text{curl } \mathbf{E} = \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{13.2c}$$

$$\text{curl } \mathbf{H} = \nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \tag{13.2d}$$

As before, we assume for the magnetic vector potential **A** to be

$$\mathbf{B} = \text{curl } \mathbf{A} = \nabla \times \mathbf{A} \tag{13.3}$$

then Eq. (13.2b) is automatically satisfied.

Substituting the value of **B** in terms of **A** in Eq. (13.2c), we get

$$\text{curl } \mathbf{E} = \nabla \times \mathbf{E} = -\left(\frac{\partial}{\partial t}\right)(\nabla \times \mathbf{A}).$$

Since the space and the time-differentiations can be reversed, i.e.

$$\nabla \times \mathbf{E} = -\left(\nabla \times \frac{\partial \mathbf{A}}{\partial t}\right)$$

which gives

$$\nabla \times \left(\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t}\right) = 0 \tag{13.48}$$

Hence $[\mathbf{E} + (\partial \mathbf{A}/\partial t)]$ is an Irrotational vector, and thus can be expressed as the gradient of a scalar, i.e. say $(= -\text{grad } V)$, where V is a scalar potential. Then

$$\mathbf{E} = -\left(\frac{\partial \mathbf{A}}{\partial t}\right) - \text{grad } V = -\left(\frac{\partial \mathbf{A}}{\partial t}\right) - \nabla V \tag{13.49}$$

Now Eqs. (13.2a) and (13.2d) remain to be dealt with, and require the assumption of the constituent relations:

$$\mathbf{D} = \varepsilon \mathbf{E} \tag{12.12}$$

$$\mathbf{B} = \mu \mathbf{H} \tag{12.13}$$

With Eq. (12.12), Eq. (13.2a) gives

$$\frac{\rho_C}{\varepsilon} = \text{div } \mathbf{E} = \nabla \cdot \mathbf{E} = \nabla \cdot \left(-\frac{\partial \mathbf{A}}{\partial t} - \nabla V\right)$$

$$= -\nabla \cdot \frac{\partial \mathbf{A}}{\partial t} - \nabla^2 \text{V} \tag{13.50}$$

and Eq. (13.2d) gives

$$\text{curl curl } \mathbf{A} = \nabla \times \nabla \times \mathbf{A} = \mu\left(\mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}\right)$$

If $\mathbf{A}$ is specified in the Cartesian coordinates, then its components are (A_x, A_y, A_z), and it may be verified that for each component of $\mathbf{A}$, $(\text{curl curl } \mathbf{A})_x = (\text{grad div } \mathbf{A})_x - \nabla^2 A_x$, and similar expressions can be found for A_y and A_z.

Hence, we can write the vector equation

$$\text{grad div } \mathbf{A} - \nabla^2 \mathbf{A} = \mu \mathbf{J} + \mu\left(\frac{\partial \mathbf{D}}{\partial t}\right)$$

$$= \mu \mathbf{J} + \mu\varepsilon\left(\frac{\partial \mathbf{E}}{\partial t}\right)$$

$$= \mu \mathbf{J} + \mu\varepsilon\left[-\frac{\partial^2 \mathbf{A}}{\partial t^2} - \text{grad}\left(\frac{\partial V}{\partial t}\right)\right]$$

by substituting from Eq. (13.49).

By rearranging the terms in the above equation:

$$\nabla^2 \mathbf{A} - \left(\frac{1}{\text{u}^2}\right)\left(\frac{\partial^2 \mathbf{A}}{\partial t^2}\right) = -\mu \mathbf{J} + \text{grad}\left[\text{div } \mathbf{A} + \left(\frac{1}{\text{u}^2}\right)\left(\frac{\partial V}{\partial t}\right)\right] \tag{13.51}$$

where $(1/\text{u}^2) = \mu\varepsilon$, so written because u has the dimension of velocity; and also:

$$\nabla^2 \mathbf{A} = \mathbf{i}_x \nabla^2 A_x + \mathbf{i}_y \nabla^2 A_y + \mathbf{i}_z \nabla^2 A_z$$

It has been remarked that **A** is not uniquely defined by Eq. (13.3), and can be made to satisfy a condition additional to:

$$\text{curl } \mathbf{A} = \nabla \times \mathbf{A} = \mathbf{B}$$

This condition we choose to be is

$$\text{div } \mathbf{A} + \left(\frac{1}{\mathrm{u}^2}\right)\left(\frac{\partial V}{\partial t}\right) = 0 \tag{13.52}$$

(this condition is known as the 'Lorentz Condition')

and then Eq. (13.51) reduces to

$$\nabla^2 \mathbf{A} - \left(\frac{1}{\mathrm{u}^2}\right)\left(\frac{\partial^2 \mathbf{A}}{\partial t^2}\right) = -\mu \mathbf{J} \tag{13.53}$$

Also by substituting for div **A** from Eq. (13.52) into Eq. (13.50), we get

$$\nabla^2 V - \left(\frac{1}{\mathrm{u}^2}\right)\left(\frac{\partial^2 V}{\partial t^2}\right) = -\frac{\rho_{\mathrm{C}}}{\varepsilon} \tag{13.54}$$

Thus by the use of the magnetic vector and the electric scalar potentials **A** and V respectively, the Maxwell's equations are reduced to two equations, i.e. (13.53) and (13.54). The vector equation (13.53) is really equivalent to three scalar equations, and this equation associates **A** with the conduction currents in the field; whilst Eq. (13.54) associates the electric scalar potential V with the charges in the system. Hence, given the system of the currents and the charges, **A** and V can be calculated; and the actual field vectors are then given by the equations:

$$\begin{aligned} \mathbf{E} &= -\left(\frac{\partial \mathbf{A}}{\partial t}\right) - \nabla V \\ \mathbf{B} &= \nabla \times \mathbf{A} \end{aligned} \tag{13.55}$$

and it is seen that the magnetic vector potential **A** is the sum of two parts; one associated with the changes in the magnetic field, and one (the electrostatic part) set up by the charges.

The condition imposed on the potentials, i.e.

$$\text{div } \mathbf{A} + \left(\frac{1}{\mathrm{u}^2}\right)\left(\frac{\partial V}{\partial t}\right) = 0 \tag{13.52}$$

is called the '*Lorentz Condition*' or the '*Lorentz Gauge*'. Such a constraint imposed on the definition of the vector potential is called a 'gauge transformation'. The advantage of this transformation is that the two potentials **A** and V have been separated into two equations which are *not* interdependent, i.e.

$$\nabla^2 \mathbf{A} - \left(\frac{1}{\mathrm{u}^2}\right)\left(\frac{\partial^2 \mathbf{A}}{\partial t^2}\right) = -\mu \mathbf{J} \tag{13.53}$$

$$\nabla^2 V - \left(\frac{1}{\mathrm{u}^2}\right)\left(\frac{\partial^2 V}{\partial t^2}\right) = -\frac{\rho_{\mathrm{C}}}{\varepsilon} \tag{13.54}$$

In this case, whilst the vector potential **A** is a consequence of the conduction currents (i.e. the changes in the magnetic field), and V is due to the charge distributions of the system, it *must* be noted that this V (though it is associated with the charges in the system) is *not* the electrostatic

potential (as defined in Chapter 1, Section 1.6.1) which was based on the potential distribution due to the static charges of the system. It (i.e. the present V based on the Lorentz gauge) is merely *a* scalar electric potential function which is a consequence of the static charges of the system.

The other condition (or the constraint):

$$\text{div } \mathbf{A} = \nabla \cdot \mathbf{A} = 0 \tag{13.4}$$

which we had imposed in the earlier sections of this chapter (i.e. Sections 13.2 to 13.4.6) is called the 'Coulomb Condition' or the 'Coulomb Gauge' and is another gauge transformation. This transformation reduces the Maxwell's equations to again two potential equations which are of a form different from the equations obtained by the Lorentz gauge. These two equations (i.e. those obtained by the Coulomb gauge) can be easily obtained from Eqs. (13.51) and (13.50) with the substitution from (13.4). The potential equations for the Coulomb condition then are

$$\nabla^2 \mathbf{A} - \left(\frac{1}{\mathrm{u}^2}\right)\left(\frac{\partial^2 \mathbf{A}}{\partial t^2}\right) = -\mu \mathbf{J} + \left(\frac{1}{\mathrm{u}^2}\right)\nabla\left(\frac{\partial V}{\partial t}\right) \tag{13.56}$$

and

$$\nabla^2 V = -\frac{\rho_{\mathrm{C}}}{\varepsilon} \tag{13.57}$$

In this case, whilst **A** and V have not been separated as before (as by the Lorentz gauge), the present scalar electric potential V is now the electrostatic potential (as defined in Section 1.6.1) of the system, based on the potential energy distribution of the system. With this transformation, the electric field vector **E** is now separable into an electrostatic field and a wave field given by $-(\partial \mathbf{A}/\partial t)$.

However in a lot of analyses, the Lorentz condition is often preferred, because it yields relativistically invariant expressions. Thus we see that the potentials are really a matter of choice for maximum convenience. It is **E** and **B** which are the physically important quantities. Maxwell's equations are invariant under the choice of the gauge.

Gauge invariance. As only the curl of **A** is fixed by the equation:

$$\mathbf{B} = \nabla \times \mathbf{A}$$

then any vector field which can be expressed as (grad f or ∇f) can be added to **A** without changing **B**. Thus if $\mathbf{A}_0$ is an acceptable vector potential, then so is

$$\mathbf{A} = \mathbf{A}_0 + (\text{grad } f) \tag{13.58}$$

Similarly if V_0 is the scalar potential corresponding to $\mathbf{A}_0$, then

$$V = V_0 - \left(\frac{\partial f}{\partial t}\right) \tag{13.59}$$

would leave **E** unchanged, and hence is acceptable. These transformations of the type given by Eqs. (13.58) and (13.59) are known as 'gauge transformations'.

Note: A point to be noted is that the Lorentz gauge (also known as the co-variant gauge)

$$\nabla \cdot \mathbf{A} + \left(\frac{1}{\mathrm{u}^2}\right)\left(\frac{\partial V}{\partial t}\right) = 0$$

is very similar in appearance to the equation of continuity which is

$$\nabla \cdot \mathbf{J} + \left(\frac{\partial \rho_C}{\partial t} \right) = 0$$

This has tempted a number of authors to say that the 'Lorentz condition' is the equation of continuity in terms of potentials, or a disguised form of the continuity equation. This is not quite a rigorously correct statement. This is because the continuity equation is based on a physical truth and holds good under all conditions. But this is not so with the Lorentz condition. This is one of the possible mathematical solutions, and is not the unique solution. If we wish to make a rigorous statement to compare these two equations, then our statement would be as given below:

> *Lorentz condition is consistent with the continuity equation because if we start by assuming the truth of the continuity equation, the truth of the Lorentz condition is not followed uniquely, i.e. it is not the only solution.*

13.6 RETARDED POTENTIALS

We have seen in Chapter 12, that in the time-varying fields the electric and the magnetic field signals propagate with the velocity of light c in free space. The classical vector analysis is concerned with the functions of the three space coordinates (x, y, z). The subjects of electrostatics and magnetostatics employ the vectors in the same way. But in the general time-varying fields, the quantities which specify a field, like the scalar potentials V, Ω and the vector potential $\mathbf{A}$ are time-variant. So these field quantities are functions of four variables (x, y, z, t). However, even in such cases, the vector analysis considers that all the quantities are taken at the same instant of time, i.e. we work with a three-dimensional section (t = constant), through the four-dimensional continuum.

Another way of taking a section of the four-space is to fix the time t at the observer field point P, and to consider all other points at 'retarded times' $\tau = (t - r/c)$ where r is the distance between the observer point P and any other point Q, and c is the velocity of light in free space (Figure 13.15). So the potential at P, due to a charge Q, varying with time, at the point Q is

$$V = \frac{Q(t)}{4\pi\varepsilon_0 r} \tag{13.60}$$

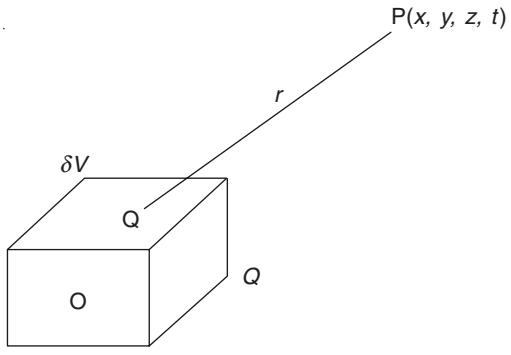


Figure 13.15 Field produced at the point P due to charges at the point Q.

at points close to the charge; and if the distances are large enough, then:

$$V = \frac{Q\left(t - \frac{r}{\mathrm{c}}\right)}{4\pi\varepsilon_0 r} \tag{13.61}$$

This potential is called the 'retarded potential' at O due to the varying charge located at the point Q. If instead of the point charge, there is a charge cloud of density $\rho_C(t)$, then the retarded scalar potential is

$$V(t) = \frac{1}{4\pi\varepsilon_0} \iiint_v \left[\frac{\rho\left(t - \frac{r}{\mathrm{c}}\right)}{r}\right] dv \tag{13.62}$$

And similarly the retarded magnetic vector potential due to a time-varying current density $\mathbf{J}(t)$ is

$$\mathbf{A}(t) = \left(\frac{\mu_0}{4\pi}\right) \iiint_v \left[\frac{\mathbf{J}\left(t - \frac{r}{\mathrm{c}}\right)}{r}\right] dv \tag{13.63}$$

The retarded values are represented by square brackets, i.e. $[\rho]$ and $[J]$. The retarded values shed new light on many aspects of electromagnetic theory, e.g. in antenna theory. Experiments show that it takes ~2.56 seconds to send radar signals to the moon and back, i.e. the signals received on the earth depend on the electric currents in the transmitter antenna at a time 2.56 seconds earlier.

The quasi-stationary electromagnetic field, where we ignore the retardation is only a limiting case. The approximation that the fields vary in synchronism with the source charges and currents is permissible only when the retardation can be justifiably ignored. This is possible at low frequencies, but not at very high frequencies.

When the charges and the currents are varying sinusoidally with time, the complex form of the retarded scalar and the vector potentials will be

$$V = \frac{1}{4\pi\varepsilon_0} \iiint_v \left[\frac{\rho_C \exp\left\{j\omega\left(t - \frac{r}{\mathrm{c}}\right)\right\}}{r}\right] dv$$

and

$$\mathbf{A}(t) = \left(\frac{\mu_0}{4\pi}\right) \iiint_v \left[\frac{\mathbf{J} \exp\left\{j\omega\left(t - \frac{r}{\mathrm{c}}\right)\right\}}{r}\right] dv \tag{13.64}$$

13.7 ELECTRIC VECTOR POTENTIAL T (WITH THE MAGNETIC SCALAR POTENTIAL Ω)

This method, which is quite analogous to that of the magnetic vector potential **A** (with the electric scalar potential V) has been comparatively recently suggested by C.J. Carpenter

(in 1977). He suggested the use of an 'electric vector potential **T**' to solve mostly eddy current problems, i.e.

$$\mathbf{J} = \nabla \times \mathbf{T} \tag{13.65}$$

This device has since then been used by a number of workers in the field of eddy current problems in electrical machines, in particular, and hence we shall present here a brief discussion of the method. Since we are restricting ourselves to eddy-current problems, we neglect here the displacement current terms, and the relevant Maxwell's equations for our consideration are:

$$\text{curl } \mathbf{H} = \nabla \times \mathbf{H} = \mathbf{J} \tag{13.2′a}$$

$$\text{curl } \mathbf{E} = \nabla \times \mathbf{E} = -\left(\frac{\partial \mathbf{B}}{\partial t}\right) \tag{13.2′b}$$

$$\text{div } \mathbf{B} = \nabla \cdot \mathbf{B} = 0 \tag{13.2′c}$$

and

$$\text{div } \mathbf{J} = \nabla \cdot \mathbf{J} = 0 \tag{13.2′d}$$

Equations (13.2′a) to (13.2′c) are the Maxwell's equations neglecting the displacement current term, and Eq. (13.2′d) is also the restricted form of the equation of continuity. The constituent relations of use to us, now, are

$$\mathbf{B} = \mu \mathbf{H} \tag{12.13}$$

$$\mathbf{J} = \sigma \mathbf{E} \tag{12.14}$$

Combining Eq. (13.65) with Eq. (13.2′a), we get

$$\nabla \times \mathbf{H} = \mathbf{J} = \nabla \times \mathbf{T}$$

or

$$\nabla \times (\mathbf{H} - \mathbf{T}) = 0$$

Hence the vector (**H** – **T**) can be expressed as the gradient of a scalar, i.e.

$$\mathbf{H} = \mathbf{T} - \nabla \Omega \tag{13.66}$$

where Ω is a magnetic scalar potential.

Note the similarity of this equation with Eq. (13.49) which states

$$\mathbf{E} = -\left(\frac{\partial \mathbf{A}}{\partial t}\right) - \nabla V \tag{13.49}$$

Now, we substitute from Eq. (13.65) into Eq. (13.2′b)

$$\nabla \times \mathbf{E} = \nabla \times \left[\left(\frac{1}{\sigma}\right) \nabla \times \mathbf{T}\right] = -\left(\frac{\partial \mathbf{B}}{\partial t}\right) = -\mu_0 \mu_r \left(\frac{\partial}{\partial t}\right)(\mathbf{T} - \nabla \Omega)$$

$$= -\mu_0 \mu_r \left(\frac{\partial \mathbf{T}}{\partial t}\right) - \nabla \left(\frac{\partial \Omega}{\partial t}\right). \tag{13.67}$$

Or, in linear media

$$\nabla \times \nabla \times \mathbf{T} = -\mu\sigma \left[\frac{\partial \mathbf{T}}{\partial t} - \nabla \left(\frac{\partial \Omega}{\partial t}\right)\right] \tag{13.68}$$

Now the operator

$$\nabla \times \nabla \times \equiv \text{grad div} - \nabla^2$$

by referring to Section 13.5 and Section 0.7.6, Eq. (0.74) (in Cartesian coordinate system).

Hence, by this substitution in Eq. (13.68) and rearranging the terms,

$$\nabla^2\mathbf{T} - \mu\sigma\left(\frac{\partial \mathbf{T}}{\partial t}\right) = \text{grad}\left[\nabla \cdot \mathbf{T} - \mu\sigma\left(\frac{\partial \Omega}{\partial t}\right)\right] \tag{13.69}$$

So far we have not defined **T** uniquely by Eq. (13.65). Hence we can impose an additional constraint, and as in the case of the magnetic vector potential, this constraint can be chosen in two possible ways, i.e. Lorentz condition which for this potential will be

$$\nabla \cdot \mathbf{T} - \mu\sigma\left(\frac{\partial \Omega}{\partial t}\right) = 0 \tag{13.70}$$

So Eq. (13.69) simplifies to

$$\nabla^2\mathbf{T} - \mu\sigma\left(\frac{\partial \mathbf{T}}{\partial t}\right) = 0 \tag{13.71}$$

The restricted form of the continuity equation (13.2d) is automatically satisfied by the potential vector **T** as

$$\text{div } \mathbf{J} = \text{div curl } \mathbf{T} = 0 \text{ is a vector identity.}$$

So now we can consider the remaining Maxwell's equation (13.2c) in which we substitute from Eq. (13.66), i.e.

$$\text{div } \mathbf{B} = \text{div } [\mu(\mathbf{T} - \nabla\Omega)] = 0$$

$$\mu\nabla \cdot \mathbf{T} - \mu\nabla \cdot \nabla\Omega = \mu\nabla^2\Omega \tag{13.72}$$

From Lorentz equation condition (13.70), the above equation reduces to

$$\nabla^2\Omega - \mu\sigma\left(\frac{\partial \Omega}{\partial t}\right) = 0 \tag{13.73}$$

Hence by using the Lorentz gauge, Maxwell's equations have been reduced to two separate equations in electric vector potential Ω, i.e.

$$\nabla^2\mathbf{T} - \mu\sigma\left(\frac{\partial \mathbf{T}}{\partial t}\right) = 0 \tag{13.74a}$$

$$\nabla^2\Omega - \mu\sigma\left(\frac{\partial \Omega}{\partial t}\right) = 0 \tag{13.74b}$$

If, on the other hand, we use Coulomb gauge for the vector potential, i.e.

$$\nabla \cdot \mathbf{T} = 0 \tag{13.75}$$

then Eq. (13.69) reduces to

$$\nabla^2\mathbf{T} - \mu\sigma\left(\frac{\partial \mathbf{T}}{\partial t}\right) = -\mu\sigma\nabla\left(\frac{\partial \Omega}{\partial t}\right) \tag{13.76a}$$

and

$$\nabla^2\Omega = 0 \tag{13.76b}$$

As in the case of the electric vector potential, **T** and Ω have not been separated, but now Ω is a Laplacian field containing no sources in the region, whereas the Lorentz gauge not only separates the two potentials, but also both satisfy the diffusion type equations.

The electric vector potential has been used to solve the problem of end region losses in turbo generators—a problem which is three dimensional in nature, and when this problem is solved by using **A** and *V*, we need to solve for all the three components of *A*, whereas using the **T**–Ω method, **T** can be simplified to produce a solution with two components of **T**. However the physical interpretations of the boundary conditions need to be handled carefully.

13.8 MAGNETIC VECTOR POTENTIAL 'A' AND THREE-DIMENSIONAL PROBLEMS

Solving electrostatic problems by using the concept of the potential (which is a scalar quantity in such problems) has been relatively easy and straightforward. We usually find the potential due to a fixed charge distribution, and then this potential is modified by superimposing a suitable perturbing potential (or potentials) to satisfy the relevant boundary conditions on dielectric or conducting boundaries of the problem. The same method is used in solving the magnetic field problems in which part of the vector potential due to the given current distribution is obtained and then it is superimposed by the suitable perturbing potentials to fulfill the requisite magnetic boundary conditions.

Though the underlying principles for both sets of the problems (i.e. electrostatics and magnetostatics, as well as electromagnetic) are the same, the problem has now become more complicated because the potential is now a vector having more than one component as distinct from the scalar potential. However in two-dimensional problems in which the vector potential has only one component the problem can be tackled by methods very similar to those of the scalar potential. But as soon as we enter the domain of three-dimensional problems in which the vector potential can have 2 or 3 components (i.e. more than one component), the method of solving the problems becomes more complicated. Even the direct method of solving the Laplace's equation (or equations with Laplacian operator) is not easy because the Laplacian operator is now being applied to a vector. This gives rise to three scalar equations, one for each component of the vector. Only in rectangular Cartesian coordinates these three components satisfy an identical equation. But then the solution of each component gets modified and altered because of different boundary conditions on the relevant boundaries of the region. This relative simplification exists in the rectangular coordinates only, because the 'unit vectors' parallel to the coordinate directions are constant both in magnitude and direction over the whole region of the problem. It should now be noted that this simplification is no longer available in other coordinate systems such as cylindrical polar and spherical polar systems. In Cartesian coordinates the Laplacian operator operates only on the magnitudes of the components of the vector potential and the unit vectors are constants with respect to the derivatives of the coordinate variables. So the effect of the Laplacian operator on all the three components of the potential vector is same. But in the other two coordinate systems, some of the unit vectors change their directions and so the Laplacian operator derivative terms have to be operated on these unit vectors producing additional non-zero terms which come out to be different for the different components of the potential vector. (This behaviour of these unit vectors has been explained and shown in Appendix 4, Section A. 4.14.). This makes the three partial differential equations for the three components of the vector potential different from one another and some of the solutions do become very complicated. Hence we have to look for a simple method, and this is done on the basis of some of physical characteristics of the magnetic field problems.

13.8.1 Orthogonal Expansions for Vector Potential (Magnetic, i.e. A)

It has been seen that in free space where there are no electric charges, the divergence of the **E** field is zero. i.e. div $\mathbf{E} = \nabla \cdot \mathbf{E} = 0$ and hence we can write $\mathbf{E} = -\text{ grad } V = -\nabla V$, where V is a scalar (potential).

Hence this **E** would satisfy the condition $\text{curl } \mathbf{E} = \nabla \times \mathbf{E} = 0$.

Similarly in free space, where there are no electric currents, the magnetic flux density **B** would have zero divergence and zero curl.

i.e. $$\nabla \cdot \mathbf{B} = 0 \qquad \text{and} \qquad \nabla \times \mathbf{B} = 0$$

In such regions, the expansions for these two fields can be written in terms of orthogonal functions in the same form. This has been done in Chapter 4, Sections. 4.2.4 to 4.2.8. The Laplace's second order partial differential equation was solved by breaking up the p.d.e. into three total o.d.e's each involving a single coordinate variable, connected by two indices. The solution thus obtained contained the two corresponding indices and six integration constants. Due to the mathematical similarity between the electric and the magnetic fields (as indicated above), it is justifiably expected that the part of the vector potential which contributes to **B** would also have the same number of indices and integration constants.

This vector potential can thus be derived from three scalar potential functions, because in the rectangular coordinates, each component of the vector potential satisfies the Laplace's equation. It should be noted that these three components are not independent but are connected by the gauge condition. We choose the Coulomb gauge here, i.e. div $\mathbf{A} = \nabla \cdot \mathbf{A} = 0$ because of its simplicity as well as its wide applicability to various eddy current problems. However this does not mean that we are restricted to eddy current problems only and the solutions obtained here apply equally well to electromagnetic radiation as well. The gauge condition implies that we are using only two independent scalar functions at most. The general expression for the vector potential, having zero divergence, is

$$\mathbf{A} = \nabla \times \mathbf{W} \tag{13.77}$$

where **W** is a vector, which can be derived from two scalar potential functions. While solving both eddy current and electromagnetic radiation problems, it is convenient to sub-divide **W** into two orthogonal components, each of which can be derived from a different scalar potential function. Thus,

$$\mathbf{W} = \mathbf{u}W_1 + \mathbf{u} \times (\nabla W_2) \tag{13.78}$$

where **u** is an arbitrary vector which has been so chosen that

$$\begin{aligned} &\nabla^2 \mathbf{A} = \nabla^2 (\nabla \times \mathbf{W}) = \nabla \times (\nabla^2 \mathbf{W}) = 0 \\ &\nabla^2 W_1 = 0 \qquad \text{and} \qquad \nabla^2 W_2 = 0 \end{aligned} \tag{13.79}$$

It can be easily verified in rectangular coordinate system, that

$$\nabla^2 (\mathbf{u}\, W_1) = \mathbf{u} \nabla^2 W_1$$

or

$$\nabla^2 (\mathbf{u}\, W_1) = \mathbf{u}\, \nabla^2 W_1 + 2 \nabla W_1 \tag{13.80}$$

and $$\nabla^2 (\mathbf{u} \times \nabla W_2) = \mathbf{u} \times \nabla (\nabla^2 W_2)$$

where $\mathbf{u} = \mathbf{i}_x, \mathbf{i}_y, \mathbf{i}_z$ or $\mathbf{i}_r$ (i.e. unit vector in the direction of r)

Because of the similarity between **B** and **E**, it can be argued that **B** (like **E**) can be derived from a single scalar function. This conclusion can be checked from the above steps of this derivation. The scalar function W_2 contributes to **A** the vector potential. Hence,

when $\mathbf{u} = \mathbf{i}_x$ or $\mathbf{i}_y$ or $\mathbf{i}_z$, $\nabla \times (\mathbf{u} \times \nabla W_2) = \mathbf{u}(\nabla^2 W_2) - \nabla(\mathbf{u} \cdot \nabla W_2)$ (13.81)

and $\mathbf{u} = \mathbf{i}_r$, $\nabla \times (\mathbf{u} \times \nabla W_2) = \mathbf{u}(\nabla^2 W_2) - \nabla(\mathbf{u} \cdot \nabla W_2) - \nabla W_2$ (13.82)

Since $\nabla^2 W_2 = 0$, the part of **A** which is derived from W_2 is then the gradient of a scalar and hence does not contribute anything to **B** in the magnetostatic problems.

(*Note:* This above derivation has been given as Problem 13.7 in this book, without the physical basis presented above and solved as Problem 9.17 in *Electromagnetism—Problems and Solutions*, 2nd Edition, 2008.)

If u_1, u_2, u_3 are orthogonal curvilinear coordinates (as explained in Appendix 4) and if **u** given above lies in the direction of u_1, and if W_1 is of the form

$$W_1 = U(u_1)\, F(u_2, u_3) \tag{13.83}$$

then $\mathbf{B} \cdot \mathbf{A} = 0$.

Next, we use this theory, so developed, to obtain the solutions of the equation

$$\nabla^2 \mathbf{A} = 0 \tag{13.84}$$

in the three different coordinate systems.

So, in terms of the general orthogonal curvilinear coordinate system, the solution would be in the form

$$\mathbf{A} = \mathbf{i}_1\, U_{11}\, U_{12}\, U_{13} + \mathbf{i}_2\, U_{21}\, U_{22}\, U_{23} + \mathbf{i}_3\, U_{31}\, U_{32}\, U_{33} \tag{13.85}$$

where $\mathbf{i}_1$, $\mathbf{i}_2$, $\mathbf{i}_3$ are the respective unit vectors in the directions of the coordinates u_1, u_2, u_3 and U_{rs} is a function of u_s only. The form of the solution should be such that the vector potential anywhere inside a specified volume, which is free of sources and is bounded by a set of surfaces, each of which is parallel to one of the coordinate surfaces, can be evaluated when the tangential component of **A** on these surfaces is given.

13.8.2 Vector Potential in Rectangular Cartesian Coordinates

In rectangular Cartesian coordinates, since each component of **A** satisfies directly the Laplace's equation in charge-free and current-free regions, there is really no need to evaluate **W** as the expressions for A_i ($i = x, y, z$) can be written down from the scalar potential solution. Hence,

$$A_i = e^{\pm\sqrt{k_x^2 + k_y^2}\, z} \begin{Bmatrix} \sin \\ \cos \end{Bmatrix} k_x x \begin{Bmatrix} \sin \\ \cos \end{Bmatrix} k_y y, \qquad i = x, y, z \tag{13.86}$$

with six constants of integration and two independent indices k_x, k_y. The solutions will be orthogonal in terms of two coordinate variables out of three. A point to be further noted is that the orthogonality of the any two coordinate variables out of the three can be suitably chosen depending on the boundary surface conditions of the parallelepiped (or cuboid) region under consideration.

13.8.3 Vector Potential in Cylindrical Polar Coordinates

The general solution of Laplace's equation in cylindrical coordinates can be built up of a sum of terms involving Bessel functions (except in some particular cases) as shown in Chapter 4, Sections 4.2.5–4.2.6. So now we consider the analogous solution for the vector potential **A** possessing orthogonal properties on the surfaces of a right circular cylinder such that the tangential components of **A** on them (i.e. on these surfaces) can be expressed as a sum of such solutions and hence determine its value at any interior point. Choosing the Bessel function solution of $\nabla^2 W_1 = 0$ in terms of ordinary Bessel function, and setting $\mathbf{u} = \mathbf{i}_z$, the **W** of Eq. (13.77) becomes

$$\mathbf{W} = \mathbf{i}_z k^{-1} (Ae^{kz} + Be^{-kz}) \{CJ_n(kr) + DY_n(kr)\} \sin(n\phi + \delta_n) \tag{13.87}$$

using simpler notations for the indices k and n instead of k_ϕ and k_z.

Deriving the vector potential from this by using Eq. (13.77),

$$\left.\begin{aligned} A_r &= -(Ae^{kz} + Be^{-kz}) \{(CJ_n(kr) + DY_n(kr)\} \frac{n}{kr} \cos(n\phi + \delta_n) \\ A_\phi &= (Ae^{kz} + Be^{-kz}) \{(CJ_n'(kr) + DY_n'(kr)\} \sin(n\phi + \delta_n) \end{aligned}\right\} \tag{13.88}$$

This is the orthogonal surface vector function. If, for a given value of z, either component of $\mathbf{i}_z \times \mathbf{A}$ vanishes for $r = a$, then at this value of z, $\mathbf{i}_z \times \mathbf{A}$ can be expressed as sum of such functions.

For suitable solutions to express the tangential component of **A** on the curved surfaces, the solution in terms of modified Bessel functions have to be used, and now the orthogonal functions are in z and ϕ. So Eq. (13.88) is now replaced by:

$$\left.\begin{aligned} A_r &= -\{(CI_n(kr) + DK_n(kr)\} \frac{n}{kr} \cos(kz + \gamma_k) \cos(n\phi + \delta_n) \\ A_\phi &= \left\{CI_n'(kr) + DK_n'(kr)\right\} \cos(kz + \gamma_k) \sin(n\phi + \delta_n) \end{aligned}\right\} \tag{13.89}$$

The z-component satisfies the scalar Laplace's equation and can be written as:

$$A_z = \left\{C'I_n(kr) + D'K_n(kr)\right\} \cos(kz + \gamma_k') \cos(n\phi + \delta_n') \tag{13.90}$$

When $k = 0$ and $n \neq 0$, the solutions for the curved surface, reduce from the above to:

$$\left.\begin{aligned} A_r &= (Ar^{n-1} + Br^{-n-1}) (Cz + D) \sin(n\phi + \delta_n) \\ A_\phi &= (-Ar^{n-1} + Br^{-n-1}) (Cz + D) \cos(n\phi + \delta_n) \\ A_z &= (A'r^n + B'r^{-n}) (C'z + D') \cos(n\phi + \delta_n') \end{aligned}\right\} \tag{13.91}$$

For the index $k = 0$, no solutions exist, which would be orthogonal in both r and ϕ and suitable for the end surfaces.

When both k and n are zero, i.e. $k = 0$, $n = 0$, some forms of interest are

$$W = \mathbf{i}_z \left[(Az\phi + B\phi + Cz + D) \ln r + (Ez + F)\phi\right] + \mathbf{r}\, Gz \tag{13.92}$$

$$\left.\begin{aligned} A_r &= Ar^{-1}z\ln r + Br^{-1}\ln r + Er^{-1}z + F \\ A_\phi &= -Ar^{-1}z\phi - Br^{-1}\phi - Cr^{-1}z - Dr^{-1} + Gr \\ A_z &= (G\phi + H)\ln r + I\phi \end{aligned}\right\} \tag{13.93}$$

13.8.4 Vector Potential in Spherical Polar Coordinates

The general solution of Laplace's equation in spherical polar coordinates has been obtained in Chapter 4, Sections 4.2.7–4.2.8 as a sum of terms involving spherical harmonics (i.e. Legendre function). Now we derive analogous solutions for the vector potential. These solutions should have orthogonal properties on the surface of a sphere so that the tangential components of the vector potential on this surface can be added up so as to obtain the value of the potential at any internal point. We choose the spherical harmonic solution of

$$\nabla^2 W_1 = 0$$

and setting $\mathbf{u} = \mathbf{i}_r$ (*Note:* $\mathbf{i}_r$ in this coordinate system is different from $\mathbf{i}_r$ of Section 13.8.3., as $\mathbf{i}_r$ in the cylindrical coordinate system is normal to the z-axis and always lies in x–y or parallel planes, whereas in this case $\mathbf{i}_r$ is in the radial direction emanating from the centre of the sphere.) So,

$$W = \mathbf{i}_r (Ar^n + Br^{-n-1})\{C P_n^m(\cos\theta) + DQ_n^m(\cos\theta)\}\sin(m\phi + \delta_m) \tag{13.94}$$

$$\left.\begin{aligned} A_\theta &= (Ar^n + Br^{-n-1})\{CP_n^m(\cos\theta) + DQ_n^m(\cos\theta)\}\, m\,\mathrm{cosec}\,\theta\cdot\cos(m\phi + \delta_m) \\ A_\phi &= (Ar^n + Br^{-n-1})\{CP_n^m(\cos\theta) + DQ_n^m(\cos\theta)\}\sin\theta\cdot\sin(m\phi + \delta_m) \end{aligned}\right\} \tag{13.95}$$

This **A** is the orthogonal surface vector function.

When $m = 0$ and $\delta_m = \pi/2$, then

$$A_\phi = (Ar^n + Br^{-n-1})\{CP_n^1(\cos\theta) + DQ_n^1(\cos\theta)\}$$

Expressing **B** in terms of **W**,

$$\mathbf{B} = \nabla\times(\nabla\times\mathbf{r}W) = -\nabla\times(\mathbf{r}\times\nabla W) = r\frac{\partial}{\partial r}(\nabla W) + 2\nabla W \tag{13.97}$$

13.8.5 Vector Potential of Eddy Currents

Induced currents in conducting media have been studied in detail in Chapter 15. (So this section will be better appreciated after studying Section 15.2, though this is not an essential requirement.) Faraday's law of electromagnetic induction, stated mathematically, gives

$$\nabla\times\mathbf{E} = -\frac{d\mathbf{B}}{dt} \tag{13.98}$$

We define the magnetic vector potential **A** as $\mathbf{B} = \nabla\times\mathbf{A}$.

Combining the above two equations, we get

$$\mathbf{E} = -\frac{d\mathbf{A}}{dt} \tag{13.99}$$

Since the electric field is produced in a conductor, there will be a current flow in it. If **J** is the current density and ρ is the resistivity of the medium (or its reciprocal conductivity = $\sigma = 1/\rho$) then, we have

$$\rho \mathbf{J} = -\frac{d\mathbf{A}}{dt} \tag{13.100}$$

The Ampere's law, neglecting the displacement current (which is justifiable here) is

$$\nabla \times \mathbf{B} = \mu \mathbf{J} \tag{13.101}$$

In terms of vector potential, this reduces to

$$\nabla^2 \mathbf{A} = -\mu \mathbf{J} \tag{13.102}$$

(by using Coulomb's gauge)

It can be easily shown by suitably combining these equations that **J**, **B** and **A** satisfy the same equation, i.e.

$$\nabla^2 \mathbf{J} = \mu\sigma \frac{d\mathbf{J}}{dt}, \quad \nabla^2 \mathbf{B} = \mu\sigma \frac{d\mathbf{B}}{dt} \quad \text{and} \quad \nabla^2 \mathbf{A} = \mu\sigma \frac{d\mathbf{A}}{dt} \tag{13.103}$$

Now as in Section 13.8.1, **A** can be written in terms of a vector **W** such that

$$\begin{aligned} \mathbf{A} &= \nabla \times \mathbf{W} \\ &= \nabla \times (\mathbf{u} W_1 + \mathbf{u} \times \nabla W_2) \end{aligned} \tag{13.104}$$

where **u** is a unit vector = $\mathbf{i}_x$ or $\mathbf{i}_y$ or $\mathbf{i}_z$ or $\mathbf{i}_r$.

As before, the vector **W** is derivable from two scalar potential functions and the two component vectors are normal to each other. Then

$$\nabla^2 \mathbf{A} = \nabla \times \{\mathbf{u}\, \nabla^2 W_1 + \mathbf{u} \times \nabla (\nabla^2 W_2)\} \tag{13.105}$$

Substituting from these two equations in Eq. (13.103),

$$\nabla \times \left[\mathbf{u} \left(\nabla^2 W_1 - \mu\sigma \frac{dW_1}{dt} \right) + \mathbf{u} \times \nabla \left(\nabla^2 W_2 - \mu\sigma \frac{dW_2}{dt} \right) \right] = 0$$

If W_1 and W_2 are the solutions of the diffusion equation, i.e.

$$\nabla^2 W = \mu\sigma \frac{dW}{dt} \tag{13.106}$$

then from these and (13.104), the vector potential of the eddy currents can be obtained. **B** can be evaluated in terms of W_1 and W_2,

$$\begin{aligned} \mathbf{B} &= \nabla \times \mathbf{A} \\ &= \nabla \times (\nabla \times \mathbf{u} W_1) + \nabla \times (\mathbf{u} \nabla^2 W_2) \end{aligned} \tag{13.107}$$

$$= \nabla \times \left[\mu\sigma \frac{d}{dt} (\mathbf{u} W_2) - \mathbf{u} \times \nabla W_1 \right] \tag{13.108}$$

where from Eqs. (13.81) and (13.82), $\mathbf{u} = \mathbf{i}_x$ or $\mathbf{i}_y$ or $\mathbf{i}_z$ or $\mathbf{i}_r$.

Since **B** and **A** satisfy the same Eq. (13.103), Eqs. (13.104) and (13.108) will be of the same form.

It should be noted that here both W_1 and W_2 contribute to **B**.

B can be further simplified to the form below by using Eqs. (13.81) and (13.82), i.e.

when $\mathbf{u} = \mathbf{i}_x$ or $\mathbf{i}_y$ or $\mathbf{i}_z$

$$\mathbf{B} = -\mu\sigma\frac{d}{dt}(\mathbf{u} \times \nabla W_2 + \mathbf{u}W_1) + \mathbf{u}\cdot\nabla(\nabla W_1) \tag{13.109}$$

and when $\mathbf{u} = \mathbf{i}_r$

$$\mathbf{B} = -\mu\sigma\frac{d}{dt}(\mathbf{i}_r \times \nabla W_2 + \mathbf{i}_r W_1) + \mathbf{i}_r \cdot \nabla(\nabla W_1) + 2\nabla W_1 \tag{13.110}$$

13.9 HERTZ VECTOR

The German physicist Heinrich Hertz, who proved experimentally the existence of electromagnetic waves in 1888 as per the theoretical predictions of Maxwell based on his study of electromagnetism, was rather uncomfortable with Maxwell's presentation of the electromagnetic equations, in particular his use of the magnetic vector potential **A**. This was because Hertz found that the vector **A** needed a scalar potential (V or ϕ) for the complete definition of the four electromagnetic equations. So he proceeded to invent a new vector quantity, which was later known as the 'Hertz Vector', such that it was totally sufficient and did not need the support of any scalar potential for defining completely these four field equations. Now, starting from the Maxwell's equations, we proceed to derive this vector.

The Maxwells equations are:

$$\text{div } \mathbf{D} = \nabla\cdot\mathbf{D} = \rho_C \tag{13.2a}$$

$$\text{div } \mathbf{B} = \nabla\cdot\mathbf{B} = 0 \tag{13.2b}$$

$$\text{curl } \mathbf{E} = \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{13.2c}$$

$$\text{curl } \mathbf{H} = \nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \tag{13.2d}$$

and the constitutive relations in linear, isotropic, homogeneous (LIH) media are

$$\mathbf{B} = \mu\mathbf{H}, \quad \mathbf{D} = \varepsilon\mathbf{E}, \quad \mathbf{E} = \rho\mathbf{J} \quad \text{or} \quad \mathbf{J} = \sigma\mathbf{E}, \quad \text{where } \sigma = 1/\rho \tag{13.2e}$$

By taking the curl of Eq. (13.2d) and combining with (13.2e),

$$\nabla \times \nabla \times \mathbf{B} = \nabla(\nabla\cdot\mathbf{B}) - \nabla^2\mathbf{B} = \mu\left\{\sigma(\nabla \times \mathbf{E}) + \varepsilon\frac{\partial}{\partial t}(\nabla \times \mathbf{E})\right\} \tag{13.111}$$

Combining Eq. (13.111) with (13.2c) and (13.2b),

$$\nabla^2\mathbf{B} = \mu\sigma\frac{\partial \mathbf{B}}{\partial t} + \mu\varepsilon\frac{\partial^2 \mathbf{B}}{\partial t^2} \tag{13.112}$$

By a similar process with Eq. (13.2c) and combining with (13.2e), (13.2d) and (13.2a), we get

$$\nabla^2 \mathbf{E} = \mu\sigma \frac{\partial \mathbf{E}}{\partial t} + \mu\varepsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} \tag{13.113}$$

provided $\rho_e = 0$ in the region, i.e. no free charge.

Both Eqs. (13.112) and (13.113) are general wave propagation equations, i.e. **B** and **E** vectors are propagating waves.

Next, we choose a general magnetic vector potential **A**, such that

$$\mathbf{B} = \text{curl } \mathbf{A} = \nabla \times \mathbf{A} \tag{13.114}$$

(*Note:* At this stage no other constraint has been imposed on **A**, i.e. neither the Coulomb gauge nor the Lorentz gauge.)

Substituting (13.113) in Eq. (13.2e),

$$\nabla \times \mathbf{E} = -\frac{\partial}{\partial t}(\nabla \times \mathbf{A}) = -\nabla \times \frac{\partial \mathbf{A}}{\partial t}$$

$$\therefore \qquad \mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} - \nabla V \tag{13.115}$$

where V is a scalar potential.

Now both **A** and V can be made to satisfy the same propagation equations [i.e. Eqs. (13.112) and (13.113)] which is a matter of convenience, but is **not** a necessary condition. We next proceed to do so. Taking curl of Eq. (13.114) and combining with (13.2d) and (13.2e),

$$\nabla \times \nabla \times \mathbf{A} = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} = \mu\sigma \mathbf{E} + \mu\varepsilon \frac{\partial \mathbf{E}}{\partial t}$$

Combining this equation with (13.115),

$$\nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} = \mu\sigma\left(-\frac{\partial \mathbf{A}}{\partial t} - \nabla V\right) + \mu\varepsilon \frac{\partial}{\partial t}\left(-\frac{\partial \mathbf{A}}{\partial t} - \nabla V\right)$$

$$= -\mu\sigma \frac{\partial \mathbf{A}}{\partial t} - \mu\varepsilon \frac{\partial^2 \mathbf{A}}{\partial t^2} - \mu\sigma \cdot \nabla V - \mu\varepsilon \nabla\left(\frac{\partial V}{\partial t}\right)$$

Rearranging the terms,

$$\nabla^2 \mathbf{A} - \mu\sigma \frac{\partial \mathbf{A}}{\partial t} - \mu\varepsilon \frac{\partial^2 \mathbf{A}}{\partial t^2} = \nabla(\nabla \cdot \mathbf{A}) + \mu\sigma \nabla V + \mu\varepsilon \nabla\left(\frac{\partial V}{\partial t}\right)$$

$$= \nabla\left(\nabla \cdot \mathbf{A} + \mu\sigma V + \mu\varepsilon \frac{\partial V}{\partial t}\right) \tag{13.116}$$

So, now we define a new vector, $\mathbf{Z}_e$, called 'Hertz electric vector', such that

$$\mathbf{A} = \mu\sigma \mathbf{Z}_e + \mu\varepsilon \frac{\partial \mathbf{Z}_e}{\partial t} \qquad \text{and} \qquad V = -\nabla \cdot \mathbf{Z}_e \tag{13.117}$$

then

$$\mu\sigma V + \mu\varepsilon \frac{\partial V}{\partial t} = -\mu\sigma \nabla \cdot \mathbf{Z}_e - \mu\varepsilon \frac{\partial}{\partial t}(\nabla \cdot \mathbf{Z}_e)$$

$$= -\nabla\cdot\left(\mu\sigma\mathbf{Z}_e + \mu\varepsilon\frac{\partial\mathbf{Z}_e}{\partial t}\right)$$

$$= -\nabla\cdot\mathbf{A} \tag{13.118}$$

using the definition of $\mathbf{A}$ and V in terms of the new vector $\mathbf{Z}_e$.

Hence, from Eq. (13.118),

$$\mu\sigma V + \mu\varepsilon\frac{\partial V}{\partial t} + \nabla\cdot\mathbf{A} = \text{bracketted term on the R.H.S. of Eq. (13.116)}$$

$$= 0$$

$\therefore$ $\mathbf{A}$ now satisfies the equation [from Eq. (13.116)]

$$\nabla^2\mathbf{A} = \mu\sigma\frac{\partial\mathbf{A}}{\partial t} + \mu\varepsilon\frac{\partial^2\mathbf{A}}{\partial t^2} \tag{13.119}$$

which is the wave propagation equation [specified in (13.112) and (13.113)].

Note:

$$\mathbf{B} = \nabla\times\mathbf{A} = \nabla\times\left(\mu\sigma\mathbf{Z}_e + \mu\varepsilon\frac{\partial\mathbf{Z}_e}{\partial t}\right)$$

$$= \mu\sigma(\nabla\times\mathbf{Z}_e) + \mu\varepsilon\frac{\partial}{\partial t}(\nabla\times\mathbf{Z}_e) \tag{13.120}$$

Next, we consider the reduction of the scalar potential to the wave propagation equation. So we start with Eqs. (13.2a) and (13.2e),

$$\nabla\cdot\mathbf{D} = \rho_C \quad \text{or} \quad \nabla\cdot\mathbf{E} = \frac{\rho_C}{\varepsilon}$$

or

$$\frac{\rho_C}{\varepsilon} = \nabla\cdot\mathbf{E} = \nabla\cdot\left(-\frac{\partial\mathbf{A}}{\partial t} - \nabla V\right), \qquad \text{from Eq. (13.115)}$$

$$= -\nabla\cdot\left(\frac{\partial\mathbf{A}}{\partial t}\right) - \nabla^2 V$$

or

$$\nabla^2 V + \frac{\rho_C}{\varepsilon} = -\frac{\partial}{\partial t}(\nabla\cdot A)$$

$$= +\frac{\partial}{\partial t}\left(\mu\sigma V + \mu\varepsilon\frac{\partial V}{\partial t}\right), \qquad \text{from Eq. (13.118).}$$

$\therefore$

$$\nabla^2 V + \frac{\rho_C}{\varepsilon} = \mu\sigma\frac{\partial V}{\partial t} + \mu\varepsilon\frac{\partial^2 V}{\partial t^2} \tag{13.121}$$

Since we are considering charge-free region, $\rho_C = 0$

$\therefore$

$$\nabla^2 V = \mu\sigma\frac{\partial V}{\partial t} + \mu\varepsilon\frac{\partial^2 V}{\partial t^2} \tag{13.122}$$

Hence, now V also satisfies the general wave propagation equation.

Thus the new vector $\mathbf{Z}_e$, called Hertz electric vector,

$$\mathbf{A} = \mu\sigma \mathbf{Z}_e + \mu\varepsilon \frac{\partial \mathbf{Z}_e}{\partial t}, \quad V = -\nabla \cdot \mathbf{Z}_e \tag{13.117}$$

completely describes the whole electromagnetic field, as it includes both $\mathbf{A}$ and V in its definition above. It should be noted that now a single vector $\mathbf{Z}_e$ has been so defined that it does what was being done by the combination of the vector potential $\mathbf{A}$ and the scalar potential V. The relationship between $\mathbf{A}$ and V from the propagation Eqs. (13.119) and (13.122) gives rise to another relation between $\mathbf{A}$ and V. This can be derived by taking the divergence of both the sides of Eq. (13.115),

i.e. $$\text{div } \mathbf{E} = \nabla \cdot \mathbf{E} = \frac{1}{\varepsilon} \nabla \cdot \mathbf{D} = -\frac{\partial}{\partial t}(\nabla \cdot \mathbf{A}) - \nabla \cdot (\nabla V)$$

or

$$\nabla^2 V = \frac{1}{\varepsilon} \rho_C - \frac{\partial}{\partial t}(\nabla \cdot \mathbf{A})$$

Since we are considering charge-free region, $\rho_C = 0$, and so

$$\nabla^2 V = -\frac{\partial}{\partial t}(\nabla \cdot \mathbf{A}) \tag{13.123}$$

Comparing this with Eq. (13.122), it is found that the above equation holds if

$$-\frac{\partial}{\partial t}(\nabla \cdot \mathbf{A}) = \mu\sigma \frac{\partial V}{\partial t} + \mu\varepsilon \frac{\partial^2 V}{\partial t^2}$$

Integrating w.r.t time, it reduces to

$$\nabla \cdot \mathbf{A} = -\mu\sigma V - \mu\varepsilon \frac{\partial V}{\partial t}$$

which is Eq. (13.118), and hence the truth of Eq. (13.123) follows.

A similar consistency for $\mathbf{A}$ for Eqs. (13.119) and (13.118) can be shown by taking the gradient of both sides of Eq. (13.118), i.e.

$$\nabla(\nabla \cdot \mathbf{A}) = -\mu\sigma \nabla V - \mu\varepsilon \frac{\partial}{\partial t}(\nabla V) \tag{13.124}$$

Its $$\text{L.H.S.} = \nabla(\nabla \cdot \mathbf{A}) = \nabla \times \nabla \times \mathbf{A} + \nabla^2 \mathbf{A}$$

$$= \nabla \times \mathbf{B} + \nabla^2 \mathbf{A}, \qquad \text{from Eq. (13.113)}$$

$$= \mu\left(\mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}\right) + \nabla^2 \mathbf{A}, \qquad \text{from Eq. (13.2d)}$$

$$= \mu\sigma \mathbf{E} + \mu\varepsilon \frac{\partial \mathbf{E}}{\partial t} + \nabla^2 \mathbf{A}, \qquad \text{from Eq. (13.2e)}$$

Its $$\text{R.H.S.} = -\mu\sigma \nabla V - \mu\varepsilon \frac{\partial}{\partial t}(\nabla V)$$

$$= \mu\sigma\left(\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t}\right) + \mu\varepsilon\left(\frac{\partial \mathbf{E}}{\partial t} + \frac{\partial^2 \mathbf{A}}{\partial t^2}\right), \qquad \text{from Eq. (13.115)}$$

Equating the L.H.S. and the R.H.S., we get

$$\nabla^2 \mathbf{A} = \mu\sigma\frac{\partial \mathbf{A}}{\partial t} + \mu\varepsilon\frac{\partial^2 \mathbf{A}}{\partial t^2}$$

which is the propagation Eq. (13.119) for **A**.

Hence the consistency is confirmed.

It can be shown that $\mathbf{Z}_e$ also satisfies the wave propagation equation, **under certain conditions**. Hence we again start from Eq. (13.115), i.e.

$$\mathbf{E} = -\nabla V - \frac{\partial \mathbf{A}}{\partial t}$$

$$= \nabla(\nabla \cdot \mathbf{Z}_e) - \mu\sigma\frac{\partial \mathbf{Z}_e}{\partial t} - \mu\varepsilon\frac{\partial^2 \mathbf{Z}_e}{\partial t^2}, \qquad \text{from Eq. (13.117)} \tag{13.125}$$

The new condition which is imposed at this stage is

$$\mathbf{E} = \nabla \times \nabla \times \mathbf{Z}_e = \nabla(\nabla \cdot \mathbf{Z}_e) - \nabla^2 \mathbf{Z}_e \tag{13.126}$$

Substituting in Eq. (13.125), we get

$$\nabla^2 \mathbf{Z}_e = \mu\sigma\frac{\partial \mathbf{Z}_e}{\partial t} + \mu\varepsilon\frac{\partial^2 \mathbf{Z}_e}{\partial t^2} \tag{13.127}$$

which again is the wave propagation equation in $\mathbf{Z}_e$.

Note that, in general terms, the Hertz vector satisfies the condition (i.e. the constraint for the uniqueness of the vector or the gauge condition),

$$\nabla \cdot \mathbf{A} + \mu\sigma V + \mu\varepsilon\frac{\partial V}{\partial t} = 0 \tag{13.118}$$

This can be shown to be a vector identity when V and **A** are substituted in terms of $\mathbf{Z}_e$ from its definition.

When there is no conducting medium, i.e. $\sigma = 0$, the above constraint reduces to

$$\nabla \cdot \mathbf{A} + \mu\varepsilon\frac{\partial V}{\partial t} = 0 \tag{13.128}$$

which is the Lorentz gauge (or condition):

Now $\mathbf{Z}_e$ is given by

$$\mathbf{A} = \mu\varepsilon\frac{\partial \mathbf{Z}_e}{\partial t} \qquad \text{and} \qquad V = -\nabla \cdot \mathbf{Z}_e \tag{13.129}$$

In this case, the Hertz vector is another potential function.

In this sense, the Hertz vector is a more general vector than the 'potential function vectors.'

We have derived the magnetic flux density vector **B** in terms of $\mathbf{Z}_e$ [i.e. Eq.(13.120)], and **E** (the electric field vector) is obtained in terms of $\mathbf{Z}_e$ from the condition imposed on it by Eq. (13.126).

Also, in non-conducting medium, from Eq. (13.115) and the constraint Eq. (13.118), eliminating V, we get

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} - \nabla V = -\frac{\partial \mathbf{A}}{\partial t} + \frac{1}{\mu\varepsilon}\int \nabla(\nabla \cdot \mathbf{A})\, dt$$

$$= -\frac{\partial \mathbf{A}'}{\partial t} \tag{13.130}$$

and $$\mathbf{B} = \nabla \times \mathbf{A}'$$

When the region is charge free (i.e. $\rho_C = 0$), the new vector potential $\mathbf{A}'$ has zero divergence [i.e. Eq. (13.2a)].

There is also a similar Hertz magnetic vector $\mathbf{Z}_m$ which has been defined by

$$V = 0 \qquad \text{and} \qquad \mathbf{A} = \nabla \times \mathbf{Z}_m \tag{13.131}$$

It can be easily seen that $\mathbf{Z}_m$ automatically satisfies the Lorentz gauge, because in this since $V = 0$, the Lorentz gauge reduces to

$$\nabla \cdot \mathbf{A} + \mu\varepsilon \frac{\partial V}{\partial t} = 0 \quad \rightarrow \quad \nabla \cdot \mathbf{A} = 0 \tag{13.132}$$

i.e. same as Coulomb gauge, and

$$\nabla \cdot \mathbf{A} = \nabla \cdot (\nabla \times \mathbf{Z}_m) = 0 \quad \text{is a vector identity.}$$

In this case too, $\mathbf{A}$ satisfies the wave propagation equation, as considering Eq. (13.116), its R.H.S. comes out to be zero, i.e.

$$\nabla \cdot \mathbf{A} + \mu\sigma V + \mu\varepsilon \frac{\partial V}{\partial t} = 0 \qquad (\because V = 0 \text{ and } \nabla \cdot \mathbf{A} = 0) \tag{13.133}$$

The expressions for $\mathbf{E}$ and $\mathbf{B}$ vectors come out to be

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} = -\frac{\partial}{\partial t}(\nabla \times \mathbf{Z}_m) \tag{13.134}$$

$$\mathbf{B} = \nabla \times \mathbf{A} = \nabla \times \nabla \times \mathbf{Z}_m$$

$$= \nabla(\nabla \cdot \mathbf{Z}_m) - \nabla^2 \mathbf{Z}_m \tag{13.135}$$

Note: Some authors have used Π_e and Π_m to denote these two vectors.

13.9.1 Some Comments on Useful Types of Vector Potential

The following comments are in fact pertinent to the analyses of antennae, or, to be more precise, of oscillating dipoles and other multipole structures which have current distributions in the source and retarded vectors have to be used.

When the current distributions in the source are given, the charge distribution can be found from the equation of continuity. As in magnetostatics and in electrostatics, the retarded vector and scalar potentials are obtained from the current and the charge respectively, by integration. The

Hertz vector, which incorporates in itself the effects of both the vector potential and the scalar potential, when taken in retarded form would give all the fields so that they satisfy the continuity equation at the source, is obtained from the current by a single integration.

These methods, in general, yield vector potentials whose divergence is not zero and have components tangential to the antenna surfaces. When all the charges are confined to perfectly conducting surfaces which bound the region of interest where the fields are to be evaluated, it is often easier and preferable to use a second type of vector potential whose divergence is zero and to eliminate the scalar potential completely as in Eq. (13.130). This vector potential is always normal to the antenna surfaces. If there is a charge outside the conductors, then $\mathbf{A}$ can be sub-divided into two parts: a solenoidal part $\mathbf{A}_1$ and a lamellar part $\mathbf{A}_2$.

$$\underset{\downarrow}{(\nabla\cdot\mathbf{A}_1=0)} \qquad\qquad \underset{\downarrow}{(\nabla\times\mathbf{A}_2=0)}$$

The time derivative of the latter can then be written as the gradient of a scalar V_2. So we start from

$$\mathbf{E}=-\frac{\partial\mathbf{A}}{\partial t}-\nabla V$$

$$\left(=-\frac{\partial\mathbf{A}'}{\partial t} \qquad (13.130)\right)$$

and taking divergence of the above

$$\nabla\cdot\mathbf{E}=-\nabla\cdot\frac{\partial\mathbf{A}_1}{\partial t}-\nabla\cdot\frac{\partial\mathbf{A}_2}{\partial t}-\nabla\cdot\nabla V$$

$$=-\nabla\cdot\frac{\partial\mathbf{A}_2}{\partial t}-\nabla^2V \qquad \left(\because\nabla\cdot\mathbf{A}_1=0\text{ — solenoidal}\right)$$

$$=-\nabla\cdot\nabla V_2-\nabla^2V$$

$$=-\nabla^2(V_2+V)=\frac{\rho}{\varepsilon} \qquad (\text{from } \nabla\cdot\mathbf{D}=\rho)$$

A new scalar ϕ can be substituted in place of $(V_2 + V)$, thus

$$\nabla^2\phi=-\frac{\rho}{\varepsilon}$$

Thus the presence of charge makes it necessary to solve the Poisson's equation for which E.S. methods are available. This is the only complication.

It should be noted that ϕ is in phase with ρ and does not involve any energy propagation.

PROBLEMS

13.1 Find the vector potential $\mathbf{A}_z$ due to two parallel infinite straight currents I flowing in the $+z$ and $-z$ directions respectively. Find the equations for the equipotentials and evaluate $\mathbf{B}$. Show that $\mathbf{B}$ satisfies the equation $\mathbf{B}\cdot\text{grad}\,A_z = 0$, and thus the lines of $\mathbf{B}$ are the curves A_z = constant.

13.2 Define the gradient of a scalar and the curl of a vector, and give their differential notations in the Cartesian coordinate system.

The scalar magnetic potential in a region is

$$\Omega = C \ln \sqrt{x^2 + y^2}$$

Find the magnetic vector potential which produces the same magnetic field, assuming that (a) the z-component of the vector potential is zero everywhere, and (b) only the z-component of the vector potential exists.

13.3 The parallel conductors intersect a plane perpendicularly at points $(-a, 0)$, $(0, 0)$ and $(a, 0)$, and carry currents of -1, $+2$, -1 units respectively. Find the equation of a line of force in terms of some parameter which distinguishes one line from another, and prove that a particular line is the rectangular hyperbole

$$2(x^2 - y^2) = a^2$$

13.4 The direction of a vector **A** is radially outwards from the origin, and its magnitude is kr^n where

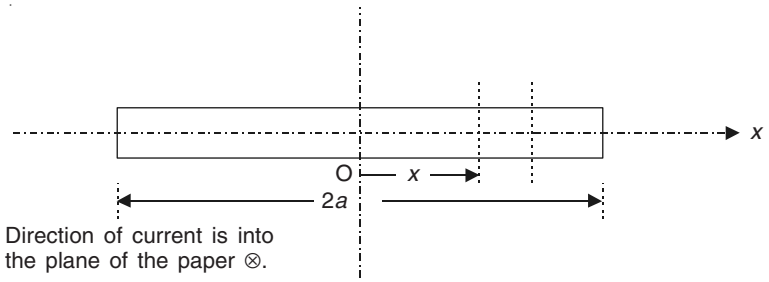
$$r^2 = x^2 + y^2 + z^2$$

Find the value of n for which div **A** = 0

Ans.: $n = -2$

13.5 A straight metal ribbon of width $2a$ and negligible thickness, carries a uniformly distributed current I. Show that the vector potential at any point on the ribbon is

$$A = -\frac{\mu_0 I}{4\pi a}\left[(a + x)\ln(a + x) + (a - x)\ln(a - x) - 2a\right]$$



Hence derive the expression for the magnetic flux density which traverses the ribbon normally to its plane at any point.

13.6 Show that the vector potential (in Cartesian coordinates) associated with a uniform magnetic field $\mathbf{B} = \mathbf{i}_z B_0$ has only two components and is given by

$$\mathbf{A} = -\mathbf{i}_x \alpha y B_0 + \mathbf{i}_y (l - \alpha) x B_0$$

where α is any arbitrary number.

Derive the expression for **A** in cylindrical coordinates and comment on the shape of lines of **A**.

13.7 The magnetic vector potential **A** is made to satisfy the constraint $\nabla \cdot \mathbf{A} = 0$. Then the general expression for the vector potential, giving zero divergence, is

$$\mathbf{A} = \nabla \times \mathbf{W} \tag{13.7.1}$$

where **W** is a vector which should be derivable from two scalar potential functions. So **W** can be split up into two orthogonal components (i.e. normal to each other), i.e.

$$\mathbf{W} = \mathbf{u}W_1 + \mathbf{u} \times \nabla W_2 \tag{13.7.2}$$

where **u** is an arbitrary vector so chosen that

$$\nabla^2\mathbf{A} = \nabla^2(\nabla \times \mathbf{W}) = \nabla \times (\nabla^2 W) = 0 \tag{13.7.3}$$

$$\nabla^2 W_1 = 0, \quad \text{and} \quad \nabla^2 W_2 = 0 \tag{13.7.4}$$

Hence, verify in rectangular Cartesian coordinates, that

$$\begin{aligned} \nabla^2\mathbf{u}W_1 &= \mathbf{u}\nabla^2 W_1 \\ &= \mathbf{u}\nabla^2 W_1 + 2\nabla W_1 \end{aligned} \tag{13.7.5}$$

and

$$\nabla^2(\mathbf{u} \times \nabla W_2) = \mathbf{u} \times \nabla(\nabla^2 W_2) \tag{13.7.6}$$

where $\mathbf{u} = \mathbf{i}_x, \mathbf{i}_y, \mathbf{i}_z$, or **r** (i.e. unit vector in the direction of **r**).

Hence for a magnetostatic field, show that the part of **A** derived from W_2 is the gradient of scalar and contributes nothing to **B**.

14 Energy Transfer in Electromagnetic Fields and Poynting Vector

14.1 INTRODUCTION

Earlier in Chapter 3, we studied the storage of potential energy in an electric field; and in Chapter 10, the storage of energy in the magnetic fields due to the motion of the charges (i.e. the current flow). We shall now consider the interaction between the electric and the magnetic energy in systems where both forms of energy are involved. Such a system can be considered analogous to mechanical systems in which both potential and kinetic energy are stored.

14.2 FLOW OF ENERGY IN THE FIELD OF A SIMPLE ELECTROMAGNETIC OSCILLATORY SYSTEM

A simple pendulum is a mechanical example of an oscillatory system in which the energy oscillates between the potential form when the bob has its maximum height, and the kinetic form when the bob has maximum velocity. Similarly a simple electromagnetic oscillatory system can be made up of an inductor connected to a capacitor in such a manner that the energy can oscillate back and forth between the two. As an example of such a system, we consider an inductor made up of a hollow metal cylinder of radius a and axial length l in free space, connected to a parallel plate capacitor having plates of length l, width d and the gap between the plates being g, as shown in Figure 14.1. When a current flows round the inductor, it charges the capacitor; and when the capacitor discharges, it causes a current to flow in the inductor. Since we are considering a low frequency operation (i.e. quasi-static), such a device will be a multi-turn one, so that its size is reasonably small. For the purpose of analysis, we shall consider a system, which is slightly modified as shown in Figure 14.2, so as to make it axially symmetrical, so that it is sufficient to take only half of the system. We also assume the metal to be a perfect conductor, and the interior of the resonator to be a vacuum. Then we can take this as a loss-less system and assuming $g << a << d$, the inductance and the capacitance to be

$$L = \left(\frac{\mu_0}{2}\right)\left(\frac{a^2}{a+d}\right), \quad C = \frac{\varepsilon_0 \pi d^2}{g} \tag{14.1}$$

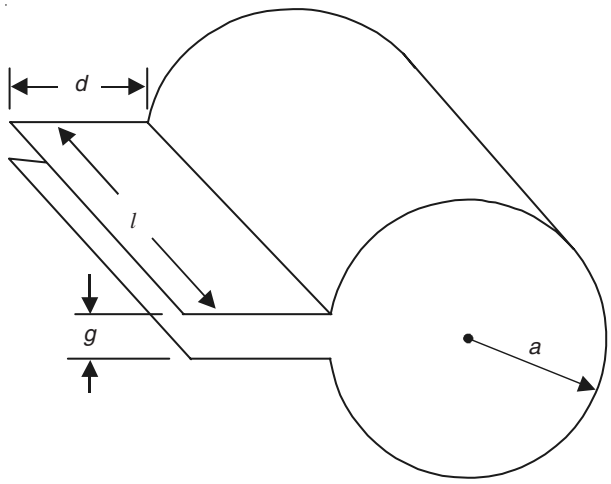


Figure 14.1 Inductor connected in parallel with a parallel plate capacitor.

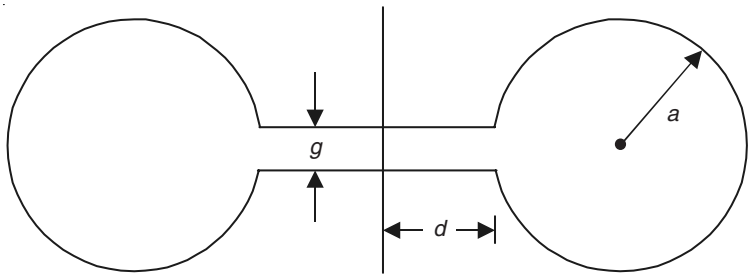


Figure 14.2 Section of a symmetrically modified system of Figure 14.1, where the inductor is a circular metal toroid and the parallel plate capacitor is now made up of circular plates of radius *d* and gap *g*.

At an instant of time t, let the current round the inductor be $I(t)$, and the voltage across the capacitor plates be $V(t)$. Let the positive direction of the current round the inductor be towards the positive plate of the capacitor, so that the **E** field in the capacitor has the same direction as the electric current, which is shown in the circuit representation of the system in Figure 14.3. At the instant of time t, the energy stored in the inductor and the capacitor are respectively

$$W_{\mathrm{m}} = \frac{1}{2} LI^2, \quad W_{\mathrm{e}} = \frac{1}{2} CV^2 \tag{14.2}$$

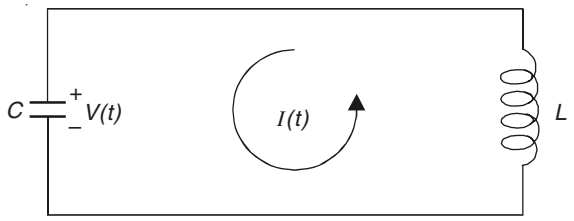


Figure 14.3 Circuit diagram of the oscillatory system.

and their time-rates of change are

$$\frac{dW_m}{dt} = LI\left(\frac{dI}{dt}\right) \quad \text{and} \quad \frac{dW_e}{dt} = CV\left(\frac{dV}{dt}\right) \tag{14.3}$$

By Faraday's law of induction,

$$V = -L\left(\frac{dI}{dt}\right) \tag{14.4}$$

and by the principle of conservation of charge,

$$I = C\left(\frac{dV}{dt}\right) \tag{14.5}$$

$\therefore$ Substituting in Eqs. (14.3), we get

$$\frac{dW_m}{dt} = -VI, \quad \frac{dW_e}{dt} = VI \tag{14.6}$$

Adding these two equations, we have

$$\frac{d}{dt}\left(W_m + W_e\right) = 0 \tag{14.7}$$

so that

$$W_m + W_e = W_0 \tag{14.8}$$

where W_0 is the total electromagnetic energy of the system which is constant. This derivation shows that the magnetic energy stored in the inductor decreases as the electric energy stored in the capacitor increases and vice versa. The total energy of the system remains constant as we have assumed it to be a loss-less one, so that the losses like radiation, conduction, and hysteresis can be neglected. Equations (14.6) show that the product VI is the rate of flow of energy from the inductor to the capacitor; and if the positive direction of V or I gets reversed, then this product would give the flow of energy from the capacitor to the inductor. If V or I is eliminated from Eqs. (14.6), then the operational equations become of the form

$$\frac{d^2I}{dt^2} + \omega^2 I = 0, \quad \frac{d^2V}{dt^2} + \omega^2 V = 0 \tag{14.9}$$

where $\omega = (LC)^{-1/2}$.

The form of the equations is that of the simple harmonic vibration, and ω is the resonant angular frequency of the system. If we measure the time from the instant that the voltage V is positive maximum V_0 at $t = 0$, then the solution of Eqs. (14.4), (14.5), or (14.9) will come out to be

$$V = V_0 \cos \omega t, \quad I = I_0 \cos\left(\omega t + \frac{\pi}{2}\right)$$

where

$$\frac{V_0}{I_0} = \omega L = \frac{1}{\omega C} = \sqrt{\frac{L}{C}} \tag{14.10}$$

Substituting from above in the energy expressions (14.2), we get

$$W_m = W_0 \sin^2 \omega t, \quad W_e = W_0 \cos^2 \omega t$$

where

$$W_0 = \frac{1}{2} LI_0^2 = \frac{1}{2} CV_0^2$$

$$= \text{ total energy of the system} \tag{14.11}$$

Hence the energy expressions can be rewritten as

$$W_m = \frac{1}{2} W_0 (1 - \cos 2\omega t), \quad W_e = \frac{1}{2} W_0 (1 + \cos 2\omega t) \tag{14.12}$$

These two equations do satisfy the energy Eq. (14.8). The complementary variations of W_m and W_e are shown in Figure 14.4. The *stored electric energy* is plotted vertically upwards from the bottom of the diagram, and the magnetic energy in the inductor is plotted vertically downwards. The period of oscillation is

$$T = 2\pi\sqrt{LC} \tag{14.13}$$

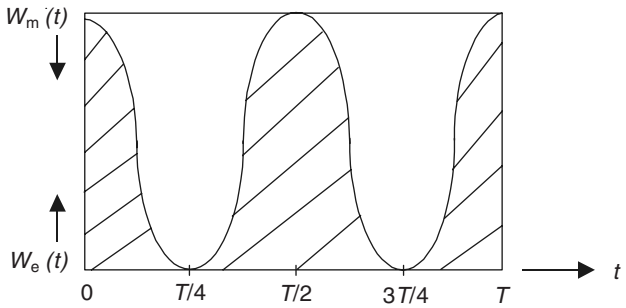


Figure 14.4 Oscillation of energy between the capacitor and the inductor of a resonant oscillatory circuit.

So it can be seen that on the *average*, half the energy of the system is stored in the inductor, and the other half in the capacitor, i.e.

$$W_{m,\,av} = W_{e,\,av} = \frac{1}{2} W_0 \tag{14.14}$$

Substituting from Eqs. (14.12) in Eq. (14.6), we get

$$VI = -\omega W_0 \sin 2\omega t = \omega W_0 \cos\left(2\omega t + \frac{\pi}{2}\right) \tag{14.15}$$

This quantity plotted as a function of time, gives the *rate of flow of energy* from the inductor to the capacitor; and when it is negative, the magnitude obtained is the rate of flow of energy from the capacitor to the inductor (Figure 14.5).

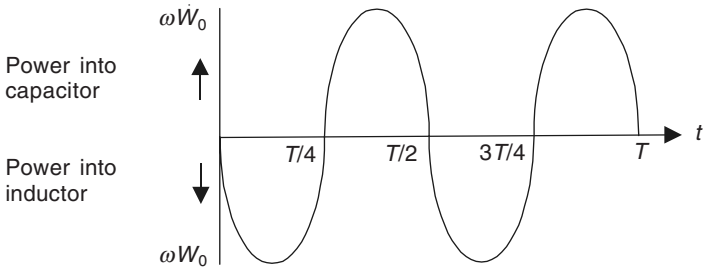


Figure 14.5 Rate of flow of energy between the inductor and the capacitor in a resonant oscillatory circuit.

14.3 FLOW OF ENERGY

We have so far considered the flow of energy in an idealized loss-less oscillatory circuit. Now we shall generalize the concept in terms of the field vectors and apply it to some more simple problems.

So we start with an electrostatic field. The energy in an electrostatic field may be considered as stored with a density

$$\int E\, dD \text{ joules/metre}^3 \tag{14.16}$$

or in a linear medium (Chapter 3, Section 3.3):

$$W_e = \frac{1}{2} DE = \frac{1}{2} \varepsilon E^2 \text{ joules/metre}^3$$
$$= \frac{1}{2} \mathbf{D} \cdot \mathbf{E} = \frac{1}{2} \varepsilon |\mathbf{E}|^2 \text{ joules/metre}^3 \tag{14.17}$$

Similarly in a magnetic field, the stored magnetic energy W_m is

$$W_m = \int HdB \text{ joules/metre}^3 \tag{14.18}$$

or in a linear medium (Chapter 11, Section 11.10):

$$W_m = \frac{1}{2} BH = \frac{1}{2} \mu H^2 \text{ joules/metre}^3$$
$$= \frac{1}{2} \mathbf{B} \cdot \mathbf{H} = \frac{1}{2} \mu |\mathbf{H}|^2 \text{ joules/metre}^3 \tag{14.19}$$

In an electromagnetic field, both the above fields are present; so in free space

$$W_e = \frac{1}{2} \varepsilon_0 E^2$$

and

$$W_m = \frac{1}{2} \mu_0 H^2 = \frac{1}{2} \mu_0 \left(\frac{B}{\mu_0} \right)^2 = \frac{1}{2} \mu_0 \left(\frac{E}{c\mu_0} \right)^2$$

[from Chapter 12, Section 12.6, Eq. (12.44)]

$$\therefore \qquad W_e = W_m = \frac{1}{2} \varepsilon_0 E^2 \tag{14.20}$$

14.3.1 Energy of a Uniform Plane Wave

We now consider the energy of the uniform plane wave described in Chapter 12, Section 12.6.

As derived and discussed in that section, the plane wave was advancing into free space in the positive z-direction of the Cartesian coordinate system. The wave-front separates a region of zero field from a region where the electric field is **E** (x-component). We now consider a small time element δt, during which the wave-front advances through a distance ($c\delta t$) (shown in Figure 14.6), and the energy ($\varepsilon_0 E^2 c\delta t$) flows from somewhere into the each square metre of the added slab. Furthermore,

$$\varepsilon_0 E^2 c = \left(\frac{\varepsilon_0 E^2}{\mu_0 \varepsilon_0} \right)^{1/2} = E^2 \left(\frac{\varepsilon_0}{\mu_0} \right)^{1/2} = EH \tag{14.21}$$

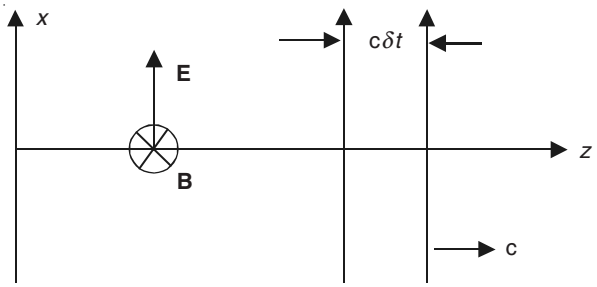


Figure 14.6 A uniform plane wave-front of **E** field travelling forward in the +*z*-direction.

We must assume that the advancing wave-front is accompanied by power EH watts/metre^2. (N.B.: E is volts/metre, H is amps/metre, and so EH is watts/metre^2.)

Since the field vectors **E** and **H** are at right angles, (EH) is the magnitude of the vector product (**E** × **H**). We have already seen that the direction of the vector (**E** × **H**) is an arrow defining the direction of propagation of the wave (see Figure 12.6 in Section 12.6 of Chapter 12). We have thus proved that in this case, the vector defined by

$$\mathbf{S} = (\mathbf{E} \times \mathbf{H}) \tag{14.22}$$

denotes the power per unit area, in both magnitude and direction. Behind the wave-front, if **E** is uniform, there is no net power in any slab; at the wave-front, the vector **S** just provides the new energy needed for the forward extension of the field.

14.3.2 Power Transmission in an Ideal Coaxial Cable

We now consider another case of power transmission. One of the simplest field configuration is given by an ideal coaxial cable (i.e. resistance-less) carrying a direct current.

The cable as shown in Figure 14.7 is resistance-less and transmits power VI from the left to the right [as per Figure 14.7(b)]. The electric and the magnetic fields are as shown, i.e. **E** is radial and **H** is circumferential [Figure 14.7(a)], so that they are mutually perpendicular everywhere. Hence (**E** × **H**) is in the direction of power flow. Also at a radius r,

$$E = \frac{Q}{2\pi\varepsilon r}$$

and so

$$V = \int_a^b \mathbf{E} \cdot d\mathbf{r} = \left(\frac{Q}{2\pi\varepsilon}\right) \ln r \Big|_a^b = \left(\frac{Q}{2\pi\varepsilon}\right) \ln\left(\frac{b}{a}\right)$$

(Q being the charge per unit length of the conductor), and so

$$E = \frac{V}{r \ln\left(\frac{b}{a}\right)} \quad \text{and} \quad H = \frac{I}{2\pi r} \tag{14.23}$$

Because of the directions of **E** and **H**, the magnitude of (**E** × **H**) is (EH). Integrating this quantity over the field space, we get

$$\int_a^b EH 2\pi r dr = \int_a^b \frac{VI\, dr}{r \ln\left(\frac{b}{a}\right)} = VI \tag{14.24}$$

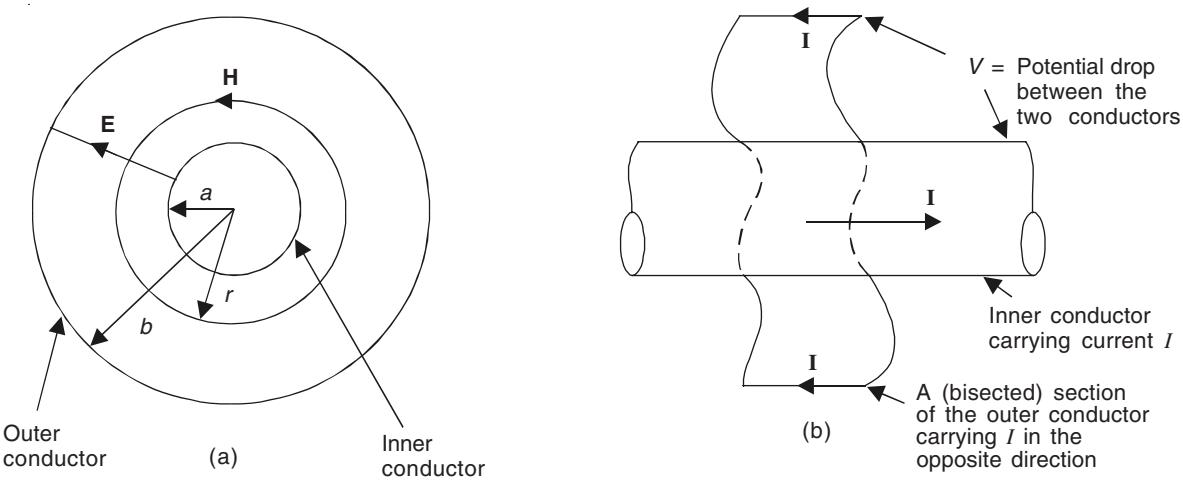


Figure 14.7 A coaxial cable carrying direct current.

This is again consistent with the hypothesis that the power is transmitted in the field with the direction and the density denoted by the vector

$$\mathbf{S} = (\mathbf{E} \times \mathbf{H}) \tag{14.25}$$

14.4 POYNTING VECTOR

We now prove in general terms that the vector **S** (called the Poynting vector) describes the energy flow. If such a vector exists, then for any closed surface Σ (Figure 14.8),

Flux of **S** into Σ = Power requirement within Σ.

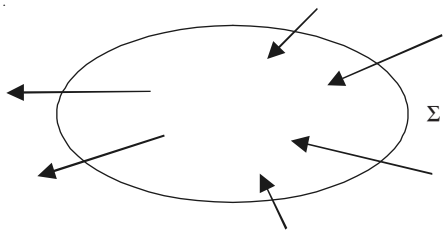


Figure 14.8 A closed surface Σ, enclosing a volume δv.

In particular, if Σ is small and encloses a volume δv, then

$$\text{Flux of } \mathbf{S} \text{ into } \Sigma = -(\text{div } \mathbf{S})\,\delta v \tag{14.26}$$

Hence, $-(\text{div } \mathbf{S})\delta v$ = Power requirement in δv

$$= \text{Power storage in } \delta v + \text{Power dissipation in } \delta v \tag{14.27}$$

Now the rate of dissipation of energy in the volume element δv, assuming an ohmic conductor, and hence the ohmic losses are

$$\frac{\partial W_1}{\partial t} = \rho J^2 \delta v$$

J being the local value of the current density vector at the instant under consideration.

But $\rho J = E$, ρ being the resistivity of the conducting material in the region. Hence,

$$\frac{\partial W_1}{\partial t} = EJ\delta v \tag{14.28}$$

Let us consider again, the rate of increase of the stored energy in the electric field within the small volume element δv—whether located in a conductor or an insulator or in free space. The value of the stored energy may be written as

$$\delta W_e = \frac{1}{2}\varepsilon_0\varepsilon_r E^2 \delta v$$

∴ In a changing electric field,

$$\text{the rate of input of energy in } \delta v = \frac{\partial W_e}{\partial t} = \frac{\partial}{\partial t}\left(\frac{1}{2}\varepsilon_0\varepsilon_r E^2\right)\delta v$$

(= power input for storage)

or

$$\frac{\partial W_e}{\partial t} = E\left(\varepsilon_0\varepsilon_r \frac{\partial E}{\partial t}\right)\delta v = \left(E\frac{\partial D}{\partial t}\right)\delta v \tag{14.29}$$

Similarly the energy input for the increase in the magnetic field in the volume element δv is

$$\delta W_m = H\delta B\delta v$$

∴ Power input for storage in the magnetic field is

$$\frac{\partial W_m}{\partial t} = \left(H\frac{\partial B}{\partial t}\right)\delta v \tag{14.30}$$

∴ The total power input to the volume for storage is

$$= \left(E\frac{\partial D}{\partial t} + H\frac{\partial B}{\partial t}\right)\delta v \tag{14.31}$$

∴ The total power requirement in the volume element δv is

$$= \left[E\left(J + \frac{\partial D}{\partial t}\right) + H\left(\frac{\partial B}{\partial t}\right)\right]\delta v \tag{14.32}$$

It should be understood that we are confining our considerations to ohmic conductors in which **E** and **J**, dielectrics in which **D** and **E**, and magnetic materials in which **H** and **B** are in the same lines. (We thus exclude iron having hysteresis—though this can also be covered in the analysis.) Under these circumstances, the right-hand side of the above Eq. (14.32) can be expressed in terms of the scalar products of the field vectors, i.e.

Power requirement in the volume element δv

$$= \left[\mathbf{E}\cdot\left(\mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}\right) + \mathbf{H}\cdot\left(\frac{\partial \mathbf{B}}{\partial t}\right)\right]\delta v$$

$$= [\mathbf{E}\cdot \text{curl } \mathbf{H} - \mathbf{H}\cdot \text{curl } \mathbf{E}]\,\delta v = [\mathbf{E}\cdot(\nabla\times\mathbf{H}) - \mathbf{H}\cdot(\nabla\times\mathbf{E})]\,\delta v \tag{14.33}$$

by substituting from the Maxwell's equations.

The bracket above is transformed by using the Cartesian components:

$$\mathbf{E}\cdot(\nabla\times\mathbf{H}) = E_x\left(\frac{\partial H_z}{\partial y}-\frac{\partial H_y}{\partial z}\right)+E_y\left(\frac{\partial H_x}{\partial z}-\frac{\partial H_z}{\partial x}\right)+E_z\left(\frac{\partial H_y}{\partial x}-\frac{\partial H_x}{\partial y}\right)$$

$$\mathbf{H}\cdot(\nabla\times\mathbf{E}) = H_x\left(\frac{\partial E_z}{\partial y}-\frac{\partial E_y}{\partial z}\right)+H_y\left(\frac{\partial E_x}{\partial z}-\frac{\partial E_z}{\partial x}\right)+H_z\left(\frac{\partial E_y}{\partial x}-\frac{\partial E_x}{\partial y}\right)$$

$\therefore$ $[\mathbf{E}\cdot(\nabla\times\mathbf{H})-\mathbf{H}\cdot(\nabla\times\mathbf{E})]$

$$=\left[-E_y\left(\frac{\partial H_z}{\partial x}\right)+E_z\left(\frac{\partial H_y}{\partial x}\right)+H_y\left(\frac{\partial E_z}{\partial x}\right)-H_z\left(\frac{\partial E_y}{\partial x}\right)\right]+\left[E_x\left(\frac{\partial H_z}{\partial y}\right)-E_z\left(\frac{\partial H_x}{\partial y}\right)-H_x\left(\frac{\partial E_z}{\partial y}\right)+H_z\left(\frac{\partial E_x}{\partial y}\right)\right]+\left[-E_x\left(\frac{\partial H_y}{\partial z}\right)+E_y\left(\frac{\partial H_x}{\partial z}\right)+H_x\left(\frac{\partial E_y}{\partial z}\right)-H_y\left(\frac{\partial E_x}{\partial z}\right)\right]$$

by rearranging and regrouping the terms.

$\therefore$ $[\mathbf{E}\cdot(\nabla\times\mathbf{H})-\mathbf{H}\cdot(\nabla\times\mathbf{E})]$

$$=-\left(\frac{\partial}{\partial x}\right)(E_yH_z-E_zH_y)-\left(\frac{\partial}{\partial y}\right)(E_zH_x-E_xH_z)-\left(\frac{\partial}{\partial z}\right)(E_xH_y-E_yH_x)$$

$$=-\text{div}\,(\mathbf{E}\times\mathbf{H})=-\nabla\cdot(\mathbf{E}\times\mathbf{H}) \tag{14.34}$$

So we have proved that the power requirement of the volume element δv is

$$-\nabla\cdot(\mathbf{E}\times\mathbf{H})\delta v \tag{14.35}$$

which is the inward flux of $\mathbf{S} = (\mathbf{E}\times\mathbf{H})$ into δv. This proves the validity of using $\mathbf{S}$ to describe the energy flow.

There are two points about this vector $\mathbf{S}$, which are worth noting:

1. $\mathbf{S}$ only denotes an energy flow if $\nabla\cdot\mathbf{S}\neq 0$. In some static fields (e.g. $\mathbf{E}$ and $\mathbf{H}$ fields of an electrically charged magnet), $\mathbf{S}$ exists, but $\nabla\cdot\mathbf{S}=0$. This $\mathbf{S}$ is meaningless (Figure 14.9).
2. There are other vectors besides $(\mathbf{E}\times\mathbf{H})$, which can represent the energy flow, for it is $\nabla\cdot\mathbf{S}$, **which is significant, and other vectors than S**, can have the same divergence at every point.

Further applications of the Poynting vector will ensue.

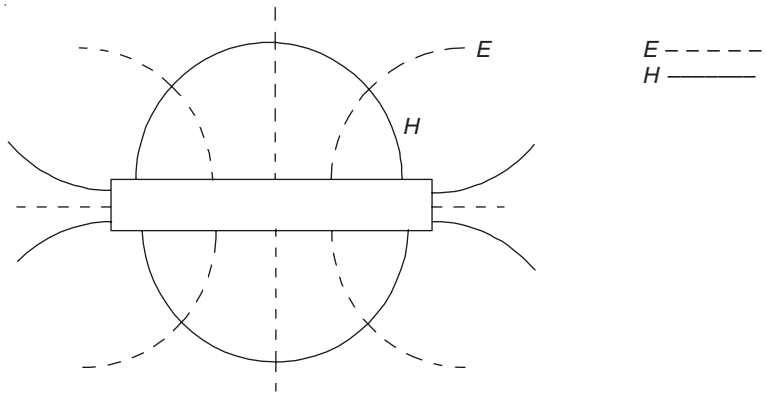


Figure 14.9 An electrically charged magnet.

14.4.1 Power Transmission in a Coaxial Cable Having Resistance, Carrying Direct Current

When the conductor resistance is taken into account, the potential difference V between the inner conductor and the outer conductor is no longer constant along the length of the cable (see Figure 14.10). However, Eqs. (14.23) still give the axial component of **S** at any point of a particular cross-section, i.e.

$$S_{\text{ax}} = E_r H_\phi = \frac{VI}{2\pi r^2} \ln\left(\frac{b}{a}\right), \tag{14.36}$$

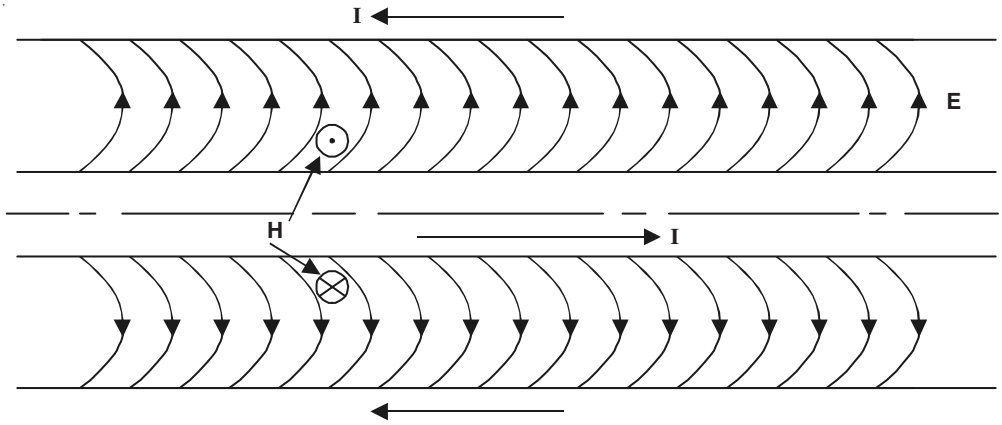


Figure 14.10 A coaxial cable having resistance and carrying a direct current.

which on integrating over the surface normal to the axis of the cable, gives the power flowing across the plane, and hence the power transmitted by the cable. (Note that the Poynting vector implies that the flow of energy is located in the field space between the conductors).

Now that the cable has a resistance as well, this would imply that another component of **E** is now present, i.e. an axial component of **E**, and hence the lines of **E** would be as shown in Figure 14.10. Let us assume that the inner conductor has a resistance of R ohms (Ω) per metre. Hence the axial component of **E** within the conductor is RI. The value of **H** on the conductor

surface is ($I/2\pi a$). These two vectors combine to give a Poynting vector, at the conductor surface which has an inward radial component, i.e.

$$S_r = -\frac{RI^2}{2\pi a} \tag{14.37}$$

This supplies the ohmic energy loss (RI^2) per unit length. Thus the flow of energy still takes place in the space between the conductors (i.e. the field space), but now a certain amount of power passes from the field to the conductors to supply the ohmic losses, so that the axial flow of power now diminishes along the length of the line. So over the length of the cable, a point will be reached where there will be no axial flow, but only a radial flow supplying the ohmic losses in the conductor.

14.4.2 Coaxial Cable (Ideal and without Resistance), Carrying a Time-varying Current

We have chosen this particular problem, because of its wide applicability. It is of use in both power engineering as well as in communications. The topic of wave transmission will be studied later in Chapter 17, though of course we have already discussed the propagation of uniform plane wave, in a highly simplified manner in Chapter 12, Section 12.6. This is another example of uniform plane wave in cylindrical polar coordinate system, which we solve from fundamentals, i.e. using Maxwell's equations in integral form instead of directly using Eq. (12.38). This problem has been chosen to show the application of Poynting vector to evaluate the energy flow of the wave, and hence the problem is solved in two stages.

A. Travelling Wave in a Coaxial Cable

We consider a coaxial cable of concentric, perfectly conducting tubes of radii a and b ($a < b$), separated by a dielectric of permittivity ε ($= \varepsilon_0\varepsilon_r$). Let a signal be applied to one end of the cable, by connecting the conductors to the terminals of a generator, whose emf is a function of time. We shall now study the behaviour of the charges, currents, and fields in the cable.

Now, if a cable is charged electrostatically, the **E** field would have been radial. So, we shall assume tentatively that the **E** field for the present cable with the time-varying charging is also radial. If the subsequent derivations are found to be consistent with all the equations, then our present assumption would be found to be justifiable. The cross-section of the cable and its axial section are shown in Figure 14.11.

To solve the problem, we consider a closed Gaussian surface Σ shown in Figure 14.11, bounded by two parallel planes z and $z + \delta z$ and a cylinder of radius r ($a < r < b$). Let the charge density on the inner conductor be Q coulombs/metre; then by Gauss' theorem, **D** at the radius r (from this Gaussian surface) is

$$D = \frac{Q\delta z}{2\pi r \delta z} = \frac{Q}{2\pi r} \tag{14.38}$$

which is same for the static case. Next, let the current in the inner conductor be I, and now we apply the Ampere's magnetic circuit law to a contour C shown in Figure 14.11(a), where C is a circle of radius r in the plane z.

The total current crossing this plane = conduction current I, as there is no contribution from the displacement current, because **D** has no component perpendicular to the plane of C.

$$\therefore \qquad \text{By Ampere's law, } H = \frac{I}{2\pi r} \tag{14.39}$$

Now, we consider a rectangular contour C_1 [shown in Figure 14.11(b)], lying in the radial plane, and having the dimensions $\delta r \times \delta z$. Applying the Maxwell's Eq. (12.10), i.e.

$$\oint_{C_1} \mathbf{E} \cdot d\mathbf{l} = \iint_{S_1} \left(\frac{d\mathbf{B}}{dt} \right) \cdot d\mathbf{S}$$

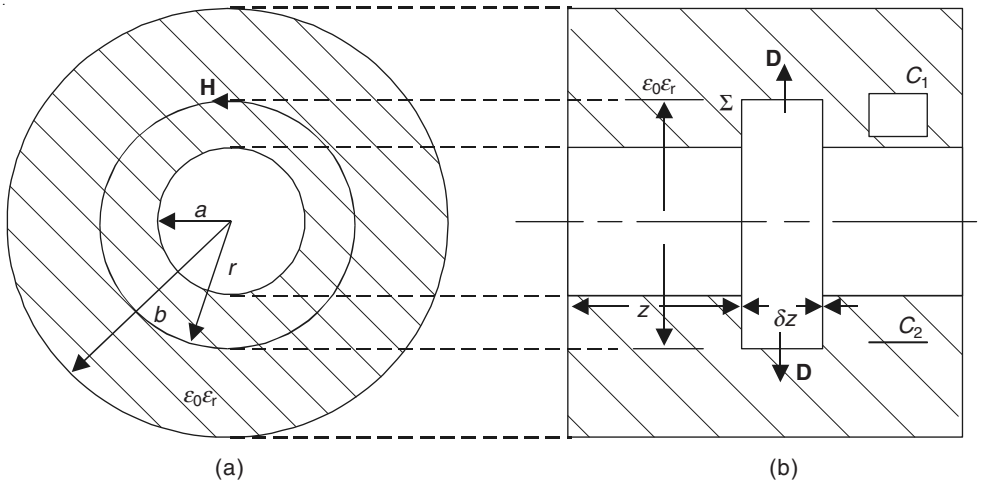


Figure 14.11 A perfect coaxial cable carrying time-varying current.

to the contour C_1, we get

$$-E\delta r + \left[E + \left(\frac{\partial E}{\partial z} \right) \delta z \right] \delta r = -\left(\frac{\partial B}{\partial t} \right) \delta z \delta r$$

or

$$\frac{\partial E}{\partial z} = -\frac{\partial B}{\partial t} \tag{14.40}$$

Similarly applying the Maxwell's Eq. (12.11) to the rectangular contour C_2 which is in the circumferential plane, we get

$$\frac{\partial H}{\partial z} = -\frac{\partial D}{\partial t} \tag{14.41}$$

The constitutive relations for this problem are

$$D = \varepsilon_0 \varepsilon_r E \qquad \text{and} \qquad B = \mu_0 H \tag{14.42}$$

Combining Eqs. (14.40) and (14.41), we obtain the same equation as for the uniform plane wave in Section 12.6, i.e.

$$\frac{\partial}{\partial z}\left[\left(\frac{1}{\varepsilon_0 \varepsilon_r} \right) \left(\frac{\partial D}{\partial z} \right) \right] = \left[\left(\frac{1}{\varepsilon_0 \varepsilon_r} \right) \left(\frac{\partial^2 D}{\partial z^2} \right) \right]$$

$$= -\left(\frac{\partial}{\partial t} \right) \left(\frac{\partial B}{\partial z} \right) = -\left(\frac{\partial}{\partial t} \right) \left(\mu_0 \frac{\partial H}{\partial z} \right)$$

$$= -\mu_0 \left(\frac{\partial}{\partial t} \right) \left(-\frac{\partial D}{\partial t} \right) = \mu_0 \left(\frac{\partial^2 D}{\partial t^2} \right)$$

or

$$\frac{\partial^2 D}{\partial z^2} = \mu_0 \varepsilon_0 \varepsilon_r \left(\frac{\partial^2 D}{\partial t^2}\right) = \left(\frac{1}{u^2}\right)\left(\frac{\partial^2 D}{\partial t^2}\right) \tag{14.43}$$

where

$u^2 = [(\mu_0 \mu_r)(\varepsilon_0 \varepsilon_r)]^{-1}$, $\mu_r = 1$ in this case; and

$c^2 = (\mu_0 \varepsilon_0)^{-1}$, and $\therefore \quad u^2 = \dfrac{c^2}{\varepsilon_r}$

From Eq. (14.38), we have

$$D = \frac{Q}{2\pi r}, \text{ where } Q \text{ varies with } z \text{ and } t.$$

$\therefore$ Substituting in Eq. (14.43), the equation becomes

$$\frac{\partial^2 Q}{\partial z^2} = \left(\frac{1}{u^2}\right)\left(\frac{\partial^2 Q}{\partial t^2}\right) \tag{14.44}$$

This equation represents, as before, a travelling wave of charge, which is moving with velocity ($\pm u$).

We now consider only the single forward travelling wave, set up in a coaxial cable of infinite length, when at the source $z = 0$, the applied potential ($= V_0$) is a function of time, i.e.

$$V_0 = F(t) \tag{14.45}$$

At any point z on the length of the cable, the potential difference ($= V$) is

$$V = \int_a^b \mathbf{E} \cdot d\mathbf{r} = \int_a^b \frac{D}{\varepsilon_0 \varepsilon_r} dr = \frac{Q}{2\pi \varepsilon_0 \varepsilon_r} \int_a^b \frac{dr}{r}$$

$$= \left(\frac{Q}{2\pi \varepsilon_0 \varepsilon_r}\right) \ln\left(\frac{b}{a}\right) \tag{14.46}$$

Now, from Section 2.8.2, we remember that the capacitance C per unit length of a coaxial cable is

$$C = \frac{2\pi \varepsilon_0 \varepsilon_r}{\ln\left(\dfrac{b}{a}\right)}, \quad \text{and hence } V = \frac{Q}{C} \tag{14.47}$$

$\therefore$ V satisfies the same equation as Q and also E.

$\therefore$ The forward travelling voltage wave can be expressed as

$$V = f(z - ut)$$

But at $z = 0$, $V = V_0 = F(t) = f(-ut)$

$\therefore$ $$f(z - ut) = F\left(t - \frac{z}{u}\right)$$

Hence the general expression for V is

$$V = F\left(t - \frac{z}{\mathrm{u}}\right)$$

$$\therefore \quad Q = CF\left(t - \frac{z}{\mathrm{u}}\right) \tag{14.48}$$

and

$$D = \frac{Q}{2\pi r} = \left(\frac{C}{2\pi r}\right)F\left(t - \frac{z}{\mathrm{u}}\right)$$

$$E = \left(\frac{C}{2\pi\varepsilon_0\varepsilon_\mathrm{r} r}\right)F\left(t - \frac{z}{\mathrm{u}}\right) \tag{14.49}$$

and from Eq. (14.41),

$$H = \left(\frac{\mathrm{u}C}{2\pi r}\right)F\left(t - \frac{z}{\mathrm{u}}\right) \quad \text{and} \quad B = \left(\frac{\mu_0 \mathrm{u}C}{2\pi r}\right)F\left(t - \frac{z}{\mathrm{u}}\right) \tag{14.50}$$

Also, since $H = \dfrac{I}{2\pi r}$, $I = \mathrm{u}CF\left(t - \dfrac{z}{\mathrm{u}}\right)$ (14.51)

Next, we consider the magnetic flux in the cable.

So, in a length δz of the cable, the total flux = $\delta z \int_a^b \mathbf{B} \cdot d\mathbf{r} = \Phi \delta z$

$$\therefore \quad \Phi\delta z = \delta z \int_a^b \frac{\mu_0 \mathrm{u}C}{2\pi r} F\left(t - \frac{z}{\mathrm{u}}\right) dr$$

$$= \delta z \left(\frac{\mu_0 \mathrm{u}C}{2\pi}\right) F\left(t - \frac{z}{\mathrm{u}}\right) \ln\left(\frac{b}{a}\right)$$

$$\therefore \quad \Phi = \left(\frac{\mu_0 \mathrm{u}C}{2\pi}\right) F\left(t - \frac{z}{\mathrm{u}}\right) \ln\left(\frac{b}{a}\right) \text{ per unit length} \tag{14.52}$$

$$\therefore \quad \frac{\Phi}{I} = \left(\frac{\mu_0}{2\pi}\right)\ln\left(\frac{b}{a}\right) = \frac{\text{inductance}}{\text{unit length}} = L$$

$$\therefore \quad LC = \left[\left(\frac{\mu_0}{2\pi}\right)\ln\left(\frac{b}{a}\right)\right]\left[\frac{2\pi\varepsilon_0\varepsilon_\mathrm{r}}{\ln\left(\frac{b}{a}\right)}\right]$$

$$= \mu_0\varepsilon_0\varepsilon_\mathrm{r} = \frac{1}{\mathrm{u}^2} \tag{14.53}$$

Also,

$$\frac{\text{P.D.}}{\text{current}} = \frac{V}{I} = \frac{F\left(t - \frac{z}{\text{u}}\right)}{\left[\text{u}CF\left(t - \frac{z}{\text{u}}\right)\right]}$$

$$= \frac{1}{\text{u}C} \text{ — measured in ohms } (\Omega)$$

Denoting this quantity by, $Z = \dfrac{1}{\text{u}C} = \dfrac{\sqrt{LC}}{C} = \sqrt{\dfrac{L}{C}}$

$$= \left(\sqrt{\frac{\mu_0}{\varepsilon_0}}\right)\left(\frac{1}{2\pi\sqrt{\varepsilon_\text{r}}}\right)\ln\left(\frac{b}{a}\right) \tag{14.54}$$

In power system electrical transients, Z is called the 'surge impedance'; and in communications, Z is called the 'characteristic impedance'. Substituting for μ_0 and ε_0, we get

$$\sqrt{\frac{\mu_0}{\varepsilon_0}} = \frac{\sqrt{4\pi \times 10^{-7}}}{8.854 \times 10^{-12}} = 377\ \Omega$$

which is the wave impedance of free space.

$$\therefore \qquad Z = 377 \times \left(\frac{1}{2\pi\sqrt{\varepsilon_0}}\right)\ln\left(\frac{b}{a}\right)\Omega \tag{14.55}$$

Thus we see that for any given values of z and t, all the electrical quantities, i.e. B, H, D, E, I are proportional. The whole pattern travels towards $z \to +\infty$ with a velocity u. Figure 14.12 explains the waves of P.D. and the total current in the cable. It shows how the conduction

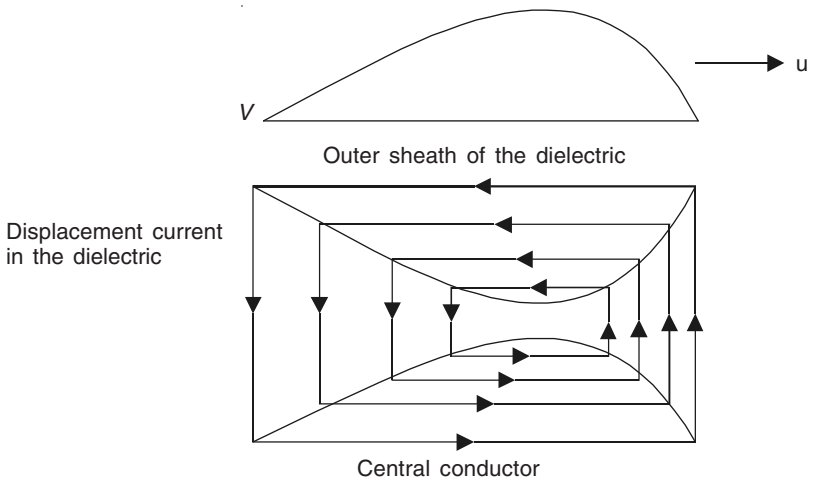


Figure 14.12 Voltage and current waves in the dielectric of the coaxial cable carrying time-varying current.

current is proportional to V and how the displacement currents form the closing links for the current paths.

B. Calculation of Power, Using the Poynting Vector

For the coaxial cable, P.D. at (z = 0) = $V_0 = F(t)$, and the electric and the magnetic fields at a section z are

$$|\mathbf{E}| = \left(\frac{C}{2\pi\varepsilon_0\varepsilon_r r}\right) F\left(t - \frac{z}{\mathrm{u}}\right) \text{—in the radial direction}$$

$$|\mathbf{H}| = \left(\frac{\mathrm{u}C}{2\pi r}\right) F\left(t - \frac{z}{\mathrm{u}}\right) \text{—in the circumferential direction}$$

$$\mathbf{S} = (\mathbf{E} \times \mathbf{H}) = \mathbf{i}_z \left(\frac{\mathrm{u}C^2}{\varepsilon_0\varepsilon_r (2\pi r)^2}\right) \left\{F\left(t - \frac{z}{\mathrm{u}}\right)\right\}^2 \tag{14.56}$$

Substituting for C from Eq. (14.47) and in terms of the surge impedance from Eq. (14.54), we get

$$\mathbf{S} = \frac{\mathbf{i}_z \left[F\left(t - \frac{z}{\mathrm{u}}\right)\right]^2}{\left(2\pi r^2\right) Z \ln\left(\frac{b}{a}\right)}$$

But, $F\left(t - \frac{z}{\mathrm{u}}\right)$ = instantaneous value of P.D. = V, and $\frac{V}{Z} = I$.

$$\therefore \quad \mathbf{S} = \frac{\mathbf{i}_z VI}{(2\pi r)^2 \left[\ln\left(\frac{b}{a}\right)\right]} \tag{14.57}$$

which is the same as the one derived for the direct current transmission. But Eq. (14.56) indicates that **S** is a travelling wave; and the direction of **S** is always the same because $[F(t - z/\mathrm{u})]^2$ can never be negative.

14.5 COMPLEX POYNTING VECTOR

All the electromagnetic quantities representing the energy or the power are quadratic functions of the field vectors and the source densities. Hence if these quantities (i.e. the field vectors and the source densities) are sinusoidal functions of time of angular frequency $\omega (= 2\pi f$, f being the time-domain frequency), then the former quantities must be expected to be the sums of two components, i.e. one independent of time and the other a sinusoidal function of frequency 2ω.

So we let $\mathbf{E}_c$ and $\mathbf{H}_c$ be the complex field vectors, i.e.

$$\mathbf{E}_c = \mathbf{E}_{Re} + j\mathbf{E}_{Im} \qquad \text{and} \qquad \mathbf{H}_c = \mathbf{H}_{Re} + j\mathbf{H}_{Im} \tag{14.58}$$

where both the real and the imaginary parts can have all the three components (x-, y-, z-), and $\mathbf{E}_{Re}$, $\mathbf{E}_{Im}$, $\mathbf{H}_{Re}$, $\mathbf{H}_{Im}$ are real quantities by themselves.

In terms of the components of these complex vectors, we can express the real sinusoidally time-varying field vectors **E** and **H** as

$$\begin{aligned}\mathbf{E} &= \text{Re}(\mathbf{E}_c e^{j\omega t}) = \text{Re}[(\mathbf{E}_{\text{Re}} + j\mathbf{E}_{\text{Im}})(\cos\ \omega t + j\ \sin\ \omega t)] \\ &= \text{Re}[(\mathbf{E}_{\text{Re}}\ \cos\ \omega t - \mathbf{E}_{\text{Im}}\ \sin\ \omega t) + j(\mathbf{E}_{\text{Re}}\ \sin\ \omega t + \mathbf{E}_{\text{Im}}\ \cos\ \omega t)] \\ &= (\mathbf{E}_{\text{Re}}\ \cos\ \omega t - \mathbf{E}_{\text{Im}}\ \sin\ \omega t)\end{aligned}$$

and similarly

$$\mathbf{H} = \mathbf{H}_{\text{Re}}\ \cos\ \omega t - \mathbf{H}_{\text{Im}}\ \sin\ \omega t \tag{14.59}$$

$\therefore$ The instantaneous value of the Poynting vector $\mathbf{S}(t)$ is given by

$S(t) = E(t)H(t)\ \sin\ \theta$ [Figure 14.13(a)]

$= (E_{\text{Re}}\ \cos\ \omega t - E_{\text{Im}}\ \sin\ \omega t)(H_{\text{Re}}\ \cos\ \omega t - H_{\text{Im}}\ \sin\ \omega t)\ \sin\ \theta$

$\therefore$ $S(t) = [E_{\text{Re}}H_{\text{Re}}\ \cos^2\omega t + E_{\text{Im}}H_{\text{Im}}\ \sin^2\omega t - (E_{\text{Re}}H_{\text{Im}} + E_{\text{Im}}H_{\text{Re}})\ \sin\ \omega t\ \cos\ \omega t)]\ \sin\ \theta$ (14.60)

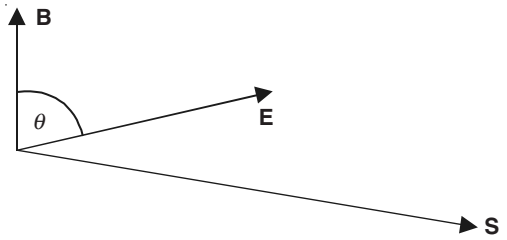


Figure 14.13(a) Poynting vector.

This instantaneous value of the Poynting vector alternates with a frequency 2ω. The average value of the Poynting vector is obtained, if the above expression is multiplied by dt, and integrated over a period of one cycle from 0 to T, where $T = 2\ \pi/\omega$, noting that over one time-period,

the average value of $\cos^2\omega t = 1/2$

the average value of $\sin^2\omega t = 1/2$

and the average value of $\sin\ \omega t\ \cos\ \omega t = 0$

$$S_{\text{av}} = \left(\frac{1}{T}\right)\int_0^T S(t)\,dt$$

$$= \frac{1}{2}(E_{\text{Re}}H_{\text{Re}} + E_{\text{Im}}H_{\text{Im}})\ \sin\ \theta \tag{14.61}$$

Expressing it as a vector, we get

$$\mathbf{S}_{\text{av}} = \frac{1}{2}(\mathbf{E}_{\text{Re}} \times \mathbf{H}_{\text{Re}} + \mathbf{E}_{\text{Im}} \times \mathbf{H}_{\text{Im}}) \tag{14.62}$$

Now we consider the vector

$$\mathbf{S}' = \mathbf{E}_c \times \mathbf{H}_c^* \tag{14.63}$$

where $\mathbf{H}_c^*$ is the complex conjugate of $\mathbf{H}_c$.

$$\therefore \quad |\mathbf{S}'| = (E_{\text{Re}} + jE_{\text{Im}})\ (H_{\text{Re}} - jH_{\text{Im}})\ \sin\ \theta$$

$$= (E_{\text{Re}}H_{\text{Re}} + E_{\text{Im}}H_{\text{Im}})\ \sin\ \theta + j(-E_{\text{Re}}H_{\text{Im}} + E_{\text{Im}}H_{\text{Re}})\ \sin\ \theta \tag{14.64}$$

$$\therefore \quad \mathbf{S}_{av} = \frac{1}{2}\mathrm{Re}(\mathbf{S}')$$

$$= \frac{1}{2}\,\mathrm{Re}(\mathbf{E}_c \times \mathbf{H}_c^*) \tag{14.64a}$$

This is the definition of the complex Poynting vector.

So expressing $\mathbf{S}$ in terms of the complex vectors,

$$\mathbf{S} = \mathbf{E} \times \mathbf{H}$$

$$= \frac{1}{2}(\mathbf{E}_c e^{j\omega t} + \mathbf{E}_c^* e^{-j\omega t}) \times \frac{1}{2}(\mathbf{H}_c e^{j\omega t} + \mathbf{H}_c^* e^{-j\omega t})$$

$$= \frac{1}{4}(\mathbf{E}_c \times \mathbf{H}_c^* + \mathbf{E}_c^* \times \mathbf{H}_c) + \frac{1}{4}(\mathbf{E}_c \times \mathbf{H}_c e^{j2\omega t} + \mathbf{E}_c^* \times \mathbf{H}_c^* e^{-j2\omega t})$$

$$= \frac{1}{2}\mathrm{Re}(\mathbf{E}_c \times \mathbf{H}_c^*) + \frac{1}{2}\mathrm{Re}(\mathbf{E}_c \times \mathbf{H}_c e^{j2\omega t}) \tag{14.65}$$

Hence Eq. (14.64a) follows directly from this equation; and it should be noted that the above equation is same as Eq. (14.60). The first term of this expression is independent of time and represents the average value of the Poynting vector. The second term is a sinusoidal vector of frequency 2ω, and its three components show the sinusoidal variations of the three components of $\mathbf{S}(t)$.

We can similarly write down the expressions for energy and power dissipation densities,

$$W_e = \frac{1}{2}\varepsilon(\mathbf{E} \cdot \mathbf{E}) = \frac{1}{4}\varepsilon(\mathbf{E}_c \cdot \mathbf{E}_c^*) + \frac{1}{4}\varepsilon\,\mathrm{Re}(\mathbf{E}_c \cdot \mathbf{E}_c e^{j2\omega t})$$

$$W_m = \frac{1}{2}\mu(\mathbf{H} \cdot \mathbf{H}) = \frac{1}{4}\mu(\mathbf{H}_c \cdot \mathbf{H}_c^*) + \frac{1}{4}\mu\,\mathrm{Re}(\mathbf{H}_c \cdot \mathbf{H}_c e^{j2\omega t})$$

$$P_D = \sigma(\mathbf{E} \cdot \mathbf{E}) = \frac{1}{2}\sigma(\mathbf{E}_c \cdot \mathbf{E}_c^*) + \frac{1}{2}\sigma\,\mathrm{Re}(\mathbf{E}_c \cdot \mathbf{E}_c e^{j2\omega t}) \tag{14.66}$$

The first terms in these three equations represent the average values of the electric energy density, of the magnetic energy density, and of the power dissipation density respectively. The second term in these equations represents the double frequency variation of the energy density, which consists of the sum of the three terms corresponding to the three components of the field vectors. They cancel out when the field vector is a vector of constant length rotating with a velocity ω as discussed in Section 0.9.4., e.g.

$$\mathbf{E}_c \cdot \mathbf{E}_c^* = |\mathbf{E}_c|^2 = 2A^2 \text{ (say)}$$

$$\mathbf{E}_c \cdot \mathbf{E}_c e^{j2\omega t} = (-A^2 + A^2)\, e^{j2\omega t} = 0$$

Another point regarding the complex Poynting vector $\mathbf{S}'$ is that whilst the real part of each component of S′ represents the average value of the corresponding component of $\mathbf{S}'$, i.e. the density of the power flow in that direction, the imaginary part of each component of $\mathbf{S}'$ may be said to represent the density of the reactive power flowing into the corresponding direction. This point is brought out by taking the divergence of the complex Poynting vector, i.e.

$$\nabla \cdot \mathbf{S}' = \nabla \cdot \left[\frac{1}{2}(\mathbf{E}_c \times \mathbf{H}_c^*)\right]$$

$$= \frac{1}{2}\left[\mathbf{H}_c^* \cdot (\nabla \times \mathbf{E}_c) - \mathbf{E}_c \cdot (\nabla \times \mathbf{H}_c^*)\right]$$

$$= -\frac{1}{2}(\mathbf{E}_c \times \mathbf{J}_c^*) - 2j\omega\left[\frac{1}{4}\mu|\mathbf{H}_c|^2 - \frac{1}{4}\varepsilon|\mathbf{E}_c|^2\right] \tag{14.67}$$

$\therefore$ $$\mathrm{Re}(\nabla \cdot \mathbf{S}') = \left(\frac{1}{2}\right)\mathrm{Re}(\mathbf{E}_c \times \mathbf{J}_c^*), \text{ and}$$

$$\mathrm{Im}(\nabla \cdot \mathbf{S}') = \left(\frac{1}{2}\right)\mathrm{Im}(\mathbf{E}_c \times \mathbf{J}_c^*) - 2\omega(W_m - W_e) \tag{14.68}$$

14.6 GENERAL COMMENTS ABOUT THE SIGNIFICANCE OF THE POYNTING VECTOR

The Poynting vector matches very well with the concept of the energy flow, in particular of wave problems (and also of various electrical machine problems). We have seen in both the wave problems studied so far, that the **E** and the **H** vectors are perpendicular to the direction of propagation of the waves, and this vector representing the energy flow is directed in the direction of propagation. In fact the Poynting vector may be regarded as an arrow, which indicates the direction of propagation at every point.

However there are problems in which difficulties arise. We have already described the difficulty with the analysis of the 'charged bar magnet' (Section 14.4, Figure 14.9). The lines of **E** and **H** vectors cross each other orthogonally in the radial planes, and thus they define a Poynting vector at right angles to the plane of Figure 14.9. This would imply that with static fields, we now have an endless source of energy. This picture is totally absurd from physical considerations. There is another problem in which a parallel plate capacitor and the poles of a U-shaped magnet are placed at right angles, as shown in Figure 14.13(b). In this problem as well, the direct application of the Poynting vector would again imply an endless source of energy. In both these cases, $\nabla \cdot \mathbf{S} = 0$, and hence even though **S** exists, it has no physical meaning.

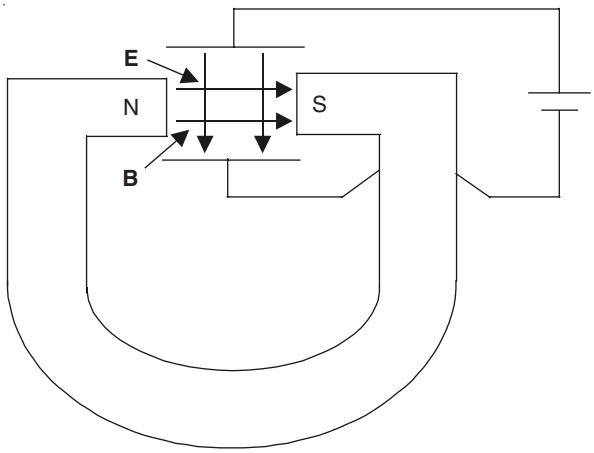


Figure 14.13(b) U-shaped magnet and parallel plate capacitor.

What should be clearly understood, is that so far we have nowhere proved that there is an energy flow of S watts/metre^2 at every point. All that has been proved is that: *the flux of* **S** *into any closed volume is = the rate of storage of energy + the rate of dissipation of energy in that volume*.

The vector **S** associated with these two problems has no flux into any of the closed volumes, and hence there is no storage or dissipation of energy taking place in these systems. Since we get this 'negative information', the vector then has no physical significance. In the time-varying problems, where there is energy transference, the vector **S** may have some meaning. But still it cannot be *definitely* said that the energy flow is equal to the vector **S**, because we have associated the energy transfer with the flux of the vector and not the vector itself. Thus there can be other vectors which can satisfy the same condition as well; as we shall see now.

14.7 ALTERNATIVE VECTORS ASSOCIATED WITH ENERGY FLOW

We now define a new vector, say **C** a function of space, such that its flux out of any volume is zero. Then, everywhere

$$\text{div } \mathbf{C} = \nabla \cdot \mathbf{C} = 0$$

Adding this **C** vectorially to the Poynting vector **S**, we get a new vector (**S** + **C**) such that

$$\nabla \cdot (\mathbf{S} + \mathbf{C}) = \nabla \cdot \mathbf{S} + \nabla \cdot \mathbf{C} = \nabla \cdot \mathbf{S} \tag{14.69}$$

$\therefore$ (**S** + **C**) would represent the flow of energy as well as **S**. Thus we find that the number of possible 'Poynting vectors' is infinite.

14.7.1 Slepian Vector

A very interesting alternative to the Poynting vector is found by taking the vector **C** as

$$\mathbf{C} = \nabla \times (V\mathbf{H}) \tag{14.70}$$

where V is the electric potential of the charges in the field. We know that

$$\nabla \cdot \nabla \times (V\mathbf{H}) = 0$$

is a vector identity, and Eq. (14.70) is a suitable choice for **C**.

Now we expand curl (V**H**), which from Section 0.7.3, Eq. (0.68), is

$$\nabla \times (V\mathbf{H}) = V(\nabla \times \mathbf{H}) - \mathbf{H} \times (\nabla V) \tag{14.71}$$

Now, we write the new vector as $\mathbf{S}_1$ which is

$$\mathbf{S}_1 = \mathbf{S} + \text{curl } (V\mathbf{H}) = (\mathbf{E} \times \mathbf{H}) + \nabla \times (V\mathbf{H})$$

For **E**, we substitute from Section 13.5, Eq. (13.49),

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} - \nabla V$$

where **A** is the magnetic vector potential.

Substituting in $\mathbf{S}_1$ from Eq. (14.71), and the above value of $\mathbf{E}$, we get

$$\begin{aligned}\mathbf{S}_1 &= \left[\left(-\frac{\partial \mathbf{A}}{\partial t} - \nabla \mathrm{V}\right) \times \mathbf{H}\right] + \mathrm{V}(\nabla \times \mathbf{H}) + [(\nabla \mathrm{V}) \times \mathbf{H}] \\ &= \mathrm{V}(\nabla \times \mathbf{H}) - \left[\left(\frac{\partial \mathbf{A}}{\partial t}\right) \times \mathbf{H}\right] \\ &= \mathrm{V}\left(\mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}\right) - \left[\left(\frac{\partial \mathbf{A}}{\partial t}\right) \times \mathbf{H}\right] \end{aligned} \tag{14.72}$$

substituting from the Maxwell's equations.

The use of the vector $\mathbf{S}_1$ to describe the flow of energy was suggested and discussed by Slepian in two papers in 1942, and this vector is known as the 'Slepian vector'. It is in fact, the sum of two vectors, the first one being V times the 'total current' and having the same direction as the current, and the second one is perpendicular to $\mathbf{H}$ and also to the part of $\mathbf{E}$, which is due to the changing magnetic field.

In fact, in these papers, Slepian discussed nine different vectors to describe the flow of energy. Before we talk about some of these other vectors, we shall apply this vector $\mathbf{S}_1$ to some of the problems already explained by the Poynting vector, and try to see the kind of picture given by this new vector.

14.7.1.1 Steady-state problems

In such problems, there are no quantities varying with time, and hence all the terms containing the operator $(\partial/\partial t)$ vanish from the equation of the vector $\mathbf{S}_1$, which then reduces to

$$\mathbf{S}_1 = V\mathbf{J} \tag{14.73}$$

This vector explains satisfactorily both the charged magnet as well as the capacitor and the U-shaped magnet problem of Section 14.6. In these problems, the static electric and magnetic fields are there, but no currents, and hence $\mathbf{S}_1 = 0$. Thus the Slepian picture gives no energy flow, which is a definite improvement on the meaningless interpretation of the Poynting vector.

We can say from this vector $\mathbf{S}_1$ of Eq. (14.73) that in all steady-state problems (i.e. where there is direct current flow), the energy flow is confined to the conducting material in the region, e.g. a current-carrying cable. Let us now consider the coaxial cable. If its resistance is negligible, V is constant throughout the length of the cable, and hence the flow across every section of the conductor is same. But when the resistance cannot be neglected, V decreases in the direction of the current flow, and the lost energy is the ohmic loss. However the distribution of energy between the 'go' and the 'return' conductor is arbitrary as the zero point of V is arbitrary.

14.7.1.2 Coaxial cable, without resistance, carrying a time-varying current

In this problem, for the Slepian vector $\mathbf{S}_1$, the terms $[V(\partial \mathbf{D}/\partial t)]$ and $[\mathbf{H} \times (\partial \mathbf{A}/\partial t)]$ have to be calculated. The directions of the field vectors are all as shown in Figure 14.14. It should be noted that $\mathbf{A}$ has the same direction as I, whether time-varying or not. For the given problem, both the vectors $[V(\partial \mathbf{D}/\partial t)]$ and $[\mathbf{H} \times (\partial \mathbf{A}/\partial t)]$ point radially outwards. Accordingly the longitudinal flow of energy is contained in the term $V\mathbf{J}$, and hence takes place only

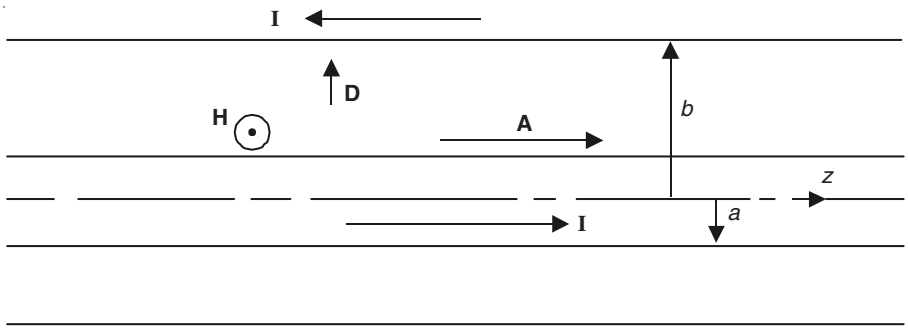


Figure 14.14 Coaxial cable, carrying time-varying current.

within the conductors. There is also a radial flow of energy in the dielectric [due to the terms $\{V(\partial \mathbf{D}/\partial t)\}$ and $\{\mathbf{H} \times (\partial \mathbf{A}/\partial t)\}$] corresponding to the storage of energy taking place there.

As shown in Figure 14.14, let the radii of the conductors be a, b where $a < b$, and we assume the potential difference between the conductors to be V, with its zero value at the radius $r_0 = \sqrt{ab}$, i.e. the geometric mean of the inner and the outer radii. Since the current is time-varying, (following the derivation of Section 14.4.2), the potential V at any radius r comes out to be

$$V = \left(\frac{C}{2\pi\varepsilon_0\varepsilon_\mathrm{r}}\right) F\left(t - \frac{z}{\mathrm{u}}\right) \ln\left(\frac{r}{r_0}\right) \tag{14.74}$$

where $V = F(t)$ at the starting point of the cable where $z = 0$. Since the knowledge of the vector potential is also required, from Eq. (13.17) we get $\mathbf{A}$ as

$$\mathbf{A} = \mathbf{A}_0 - \mathbf{i}_z \left(\frac{\mu_0 I}{2\pi}\right) \ln r$$

where the constant of integration has to be adjusted to make $\mathbf{A} = 0$ at $r = r_0$, and substituting the value of I as it is time-varying now,

$$\mathbf{A} = -\mathbf{i}_z \left(\frac{\mu_0 \mathrm{u} C}{2\pi}\right) F\left(t - \frac{z}{\mathrm{u}}\right) \ln\left(\frac{r}{r_0}\right) \tag{14.75}$$

To evaluate the radial component of $\mathbf{S}_1$, we get

$$V\left(\frac{\partial \mathbf{D}}{\partial t}\right) = \mathbf{i}_r \left[\left(-\frac{C}{2\pi\varepsilon_0\varepsilon_\mathrm{r}}\right) F\left(t - \frac{z}{\mathrm{u}}\right) \ln\left(\frac{r}{r_0}\right)\right]\left[\left(\frac{C}{2\pi r}\right) F'\left(t - \frac{z}{\mathrm{u}}\right)\right]$$

$$= \mathbf{i}_r \left(\frac{C^2}{4\pi^2\varepsilon_0\varepsilon_\mathrm{r}}\right) F\left(t - \frac{z}{\mathrm{u}}\right) F'\left(t - \frac{z}{\mathrm{u}}\right)\left(\frac{l}{r}\right) \ln\left(\frac{r}{r_0}\right) \tag{14.76}$$

from Eqs. (14.74) and (14.49).

The second term $[\mathbf{H} \times (\partial \mathbf{A}/\partial t)]$ is obtained from Eqs. (14.75) and (14.50) as

$$\mathbf{H} \times \frac{\partial \mathbf{A}}{\partial t} = \mathbf{i}_r \left[\left(\frac{uC}{2\pi r}\right) F\left(t - \frac{z}{\mathrm{u}}\right)\right]\left[\left(-\frac{\mu_0 \mathrm{u} C}{2\pi}\right) F'\left(t - \frac{z}{\mathrm{u}}\right) \ln\left(\frac{r}{r_0}\right)\right]$$

$$= -\mathbf{i}_r \left(\frac{\mu_0 \mathrm{u}^2 C^2}{4\pi^2} \right) F\left(t - \frac{z}{\mathrm{u}} \right) F'\left(t - \frac{z}{\mathrm{u}} \right) \left(\frac{1}{r} \right) \ln\left(\frac{r}{r_0} \right)$$

$$= -\mathbf{i}_r \left(\frac{C^2}{4\pi^2 \varepsilon_0 \varepsilon_\mathrm{r}} \right) F\left(t - \frac{z}{\mathrm{u}} \right) F'\left(t - \frac{z}{\mathrm{u}} \right) \left(\frac{1}{r} \right) \ln\left(\frac{r}{r_0} \right) \tag{14.77}$$

as $\mathrm{u}^2 = 1/(\mu_0 \varepsilon_0 \varepsilon_\mathrm{r})$.

These two contributions are equal, and hence the Slepian vector $\mathbf{S}_1$ for this problem is

$$\mathbf{S}_1 = V\left(J + \frac{\partial \mathbf{D}}{\partial t} \right) - \left(\frac{\partial \mathbf{A}}{\partial t} \times \mathbf{H} \right)$$

$$= \mathbf{i}_z \mathrm{u} CV\, F\left(t - \frac{z}{\mathrm{u}} \right) - \mathbf{i}_r \left(\frac{C^2}{2\pi^2 \varepsilon_0 \varepsilon_\mathrm{r}} \right) F\left(t - \frac{z}{\mathrm{u}} \right) F'\left(t - \frac{z}{\mathrm{u}} \right) \left(\frac{1}{r} \right) \ln\left(\frac{r}{r_0} \right) \tag{14.78}$$

The second term in the above expression gives a radial flow of power such that (the functions F and F' being positive) the radial energy flow is outwards when r lies between a and r_0, and inwards when r lies between r_0 and b. The electric and the magnetic energy in the dielectric appear to be as being supplied by a radial flow from the conductors. So for such a line, we can connect two wattmeters at different points on the line. Since a wattmeter measures the product of the current through the line with the P.D. between the conductors, the difference in the readings of the two meters would give a measure of the outflow of the energy from the cable to the dielectric in the intervening length of the cable.

Thus it would seem that the picture given by this Slepian vector $\mathbf{S}_1$ fits well with our visualization derived from physical considerations. But the Poynting vector is algebraically simpler. So we use both the concepts as complementary to solve our problems, and use one, which is more convenient for a particular problem.

14.7.2 Alternative Slepian Vectors for Energy Flow

As pointed out earlier in Section 14.7, the Poynting vector **S** has not been uniquely defined, since it is the divergence of **S**, which has the physical significance. We can thus add to **S** another vector, which could still represent the energy flow so long as the divergence of **S** does not change. As an example, we derived the first Slepian vector $\mathbf{S}_1$ [Eq. (14.72)], which was then applied to some practical problems (Sections 14.7.1.1 to 14.7.1.2) to check the efficacy of the new vector. We have also stated that in the papers mentioned earlier, Slepian derived nine such vectors. We shall indicate some of these now, without going into the detailed derivation. Readers, who are further interested in the subject, are strongly recommended the papers by Slepian, in particular the paper published in the *Journal of Applied Physics*, August, 1942.

The following two vectors are obvious extensions from the Poynting vector, i.e.

$$\mathbf{S}_2 = [\mathbf{E} \times \mathbf{H} - \boldsymbol{\nabla} \times (M\mathbf{E})] \tag{14.79}$$

$$\mathbf{S}_3 = [\mathbf{E} \times \mathbf{H} - \boldsymbol{\nabla} \times (N\mathbf{H})] \tag{14.80}$$

where M and N are scalar functions of position, e.g. a, b, $(\mathbf{E} \cdot \mathbf{H})$, etc. Slepian related the scalar M, N with the 'magnetic shell' which he created for the purpose of distinguishing between the

displacement and the conduction current. He argued that the shell because of its infinitesimal thickness would have negligible amount of energy in its volume; but discontinuities would be produced in the field vectors, and hence a surface distribution of energy has to be assigned to these discontinuities. The energy in the shell (say, for inclusion of infinitely permeable material) was produced by bringing in the 'compensating charges', presumably from points at infinity, and this would need the storage of certain amount of energy ascribed to the shell. But this would not interfere with the system under consideration (according to Slepian), because the system behaviour was being studied after the shell had already been created.

Another vector suggested was:

$$\mathbf{S}_4 = \left[(\mathbf{E} \times \mathbf{H}) + \frac{\partial}{\partial t} (\mathbf{A} \times \mathbf{H}) \right]$$

$$= \left[\left(\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} \right) \times \mathbf{H} + \left(\mathbf{A} \times \frac{\partial \mathbf{H}}{\partial t} \right) \right] \tag{14.81}$$

It should be noted that the divergence of this vector is different from that of the Poynting vector, the difference in energy being attributed to the compensating charges,

$$\begin{aligned} \nabla \cdot \mathbf{S}_4 - \nabla \cdot \mathbf{S} &= \nabla \cdot \left[\frac{\partial}{\partial t} (\mathbf{A} \times \mathbf{H}) \right] \\ &= \nabla \cdot \left[\frac{\partial \mathbf{A}}{\partial t} \times \mathbf{H} + \mathbf{A} \times \frac{\partial \mathbf{H}}{\partial t} \right] \\ &= \mathbf{H} \cdot \left(\nabla \times \frac{\partial \mathbf{A}}{\partial t} \right) - \left(\frac{\partial \mathbf{A}}{\partial t} \right) \cdot (\nabla \times \mathbf{H}) + \left(\frac{\partial \mathbf{H}}{\partial t} \right) \cdot (\nabla \times \mathbf{A}) - \mathbf{A} \cdot \left(\nabla \times \frac{\partial \mathbf{H}}{\partial t} \right) \\ &= \left(\frac{\partial}{\partial t} \right) \left[\mathbf{H} \cdot (\nabla \times \mathbf{A}) - \mathbf{A} \cdot \left(\mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \right) \right] \end{aligned} \tag{14.82}$$

the second term being the energy associated with the compensating charges.

In Eq. (14.81), it is possible to substitute

$$\left(\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} \right) = -\nabla V$$

and another form of the Slepian vector is obtained from $\mathbf{S}_4$, i.e.

$$\mathbf{S}_5 = -(\nabla V \times \mathbf{H}) + \left(\mathbf{A} \times \frac{\partial \mathbf{H}}{\partial t} \right)$$

$$= -\left[\nabla \times (V\mathbf{H}) \right] + V(\nabla \times \mathbf{H}) + \left(\mathbf{A} \times \frac{\partial \mathbf{H}}{\partial t} \right) \tag{14.83}$$

$$= -\left[\nabla \times (V\mathbf{H}) \right] + V\left(\mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \right) + \left(\mathbf{A} \times \frac{\partial \mathbf{H}}{\partial t} \right) \tag{14.84}$$

As far as its divergence is concerned, (14.84) is equivalent to

$$\mathbf{S}_6 = V\left(\mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \right) + \left(\mathbf{A} \times \frac{\partial \mathbf{H}}{\partial t} \right) \tag{14.85}$$

Another alternate energy vector suggested was:

$$\mathbf{S}_7 = (\mathbf{E} \times \mathbf{H}) - \frac{\partial}{\partial t}(V\mathbf{D}) \tag{14.86}$$

$$= -\left[\nabla \times (V\mathbf{H})\right] + V\mathbf{J} - \left(\frac{\partial V}{\partial t}\right)\mathbf{D} - \left(\frac{\partial \mathbf{A}}{\partial t} \times \mathbf{H}\right) \tag{14.87}$$

This vector has the same divergence as the following vector suggested by Slepian:

$$\mathbf{S}_8 = V\mathbf{J} - \left(\frac{\partial V}{\partial t}\right)\mathbf{D} - \left(\frac{\partial \mathbf{A}}{\partial t} \times \mathbf{H}\right) \tag{14.88}$$

However $\mathbf{S}_1$, suggested in Section 14.7.1 with the same divergence as the Poynting vector, is the most commonly used one.

14.8 MACDONALD VECTOR

Apart from the Slepian vectors, a number of other vectors have been suggested on the grounds of various objections raised against the Poynting vector. However the Poynting vector has the great advantage of simplicity and has gained nearly universal acceptance. Sumpner however discussed the energy flow on the basis of Huygen's principle and concluded that the Poynting vector did not describe the actual movement of energy.

A more serious objection to the Poynting's theorem was made by Macdonald in 1902, when he pointed out that the expression for the magnetic field energy in a closed volume had been derived by Maxwell for static magnetic field only. Also, Maxwell's volume integration had to be carried out over all space. Macdonald suggested, as the correct energy flow vector, the expression

$$(\mathbf{E} \times \mathbf{H}) + \frac{1}{2}\left[\left(\frac{\partial \mathbf{A}}{\partial t}\right) \times \mathbf{H} + \mathbf{A} \times \left(\frac{\partial \mathbf{H}}{\partial t}\right)\right] \tag{14.89}$$

which is very similar to $\mathbf{S}_4$ of Slepian vectors. $\mathbf{A}$ is the magnetic vector potential or as Macdonald calls it the 'Electrokinetic Momentum'. However Macdonald also overlooked the fact that the Poynting's expression for the electric field energy was open to the same objections as those for the magnetic field energy.

14.9 SILLARS' VECTOR

This was suggested to Prof. Hammond by Dr. Sillars. At low frequencies, the magnetic field $\mathbf{H}$ can be approximately represented by a scalar potential Ω, i.e.

$$\mathbf{H} = -\nabla\Omega$$

$$\therefore \quad \int_S (\mathbf{E} \times \mathbf{H}) \cdot d\mathbf{S} = \int_S (-\mathbf{E} \times \nabla\Omega) \cdot d\mathbf{S}$$

$$= \int_S \left[\nabla \times (\Omega\mathbf{E}) - \Omega\,(\nabla \times \mathbf{E})\right] \cdot d\mathbf{S}$$

$$= \int_S -\Omega\,(\nabla \times \mathbf{E}) \cdot d\mathbf{S}$$

$$= \int_S \Omega\left(\frac{\partial \mathbf{B}}{\partial t}\right) \cdot d\mathbf{S}$$

∴ The alternative vector is

$$\mathbf{S}'' = \Omega\left(\frac{\partial \mathbf{B}}{\partial t}\right) \tag{14.90}$$

This expression is also applicable to high frequency problems.

14.10 GENERAL COMMENTS

It would appear that the Poynting's theorem does not achieve its aim, i.e. it does not give the insight, as hoped by its author, into the mechanism of energy flow in space. *But it does provide a simple method of calculating the open-circuit energy flow*. As we proceed with our study of electromagnetic problems, we shall show wherever possible the use of the Poynting vector for evaluating the energy flow for both the low-frequency as well as the high-frequency problems. The wide use of the Poynting vector in high-frequency problems is quite well known. That it has also been used extensively for the analysis of low-frequency electrical machines is not so well known, and we hope to emphasize this point in our subsequent study.

PROBLEMS

14.1 A long straight uniform wire carries a steady current I. If the potential difference across a length l is V, find the value of the Poynting vector at a distance r from the wire. Hence show that the energy flowing into the wire is VI per unit time.

14.2 The validity of the Poynting vector **S** as a measure of energy flow is based on the equation:

Power requirement of a region = Inward flux of **S** into the region.

If the region under consideration is a slice of coaxial cable carrying a direct current I, then this equation reduces to 0 = 0, and yet the integral of **S** across the cross-section of the dielectric correctly gives the power flow into the cable. Explain.

14.3 Prove that the Poynting vector in a uniform plane electromagnetic wave in vacuo has a magnitude $E^2/(\mu_0 C)$.

14.4 The radiation energy of the sun, received on the earth's surface has the mean value of 1.54 kW/m^2, and is normal to the direction of the rays. Find the rms values of **E** and **H** in a polarized light beam having the same energy flow as the sunlight.

Ans.: E = 761.6 V/m H = 2.02 amps/m

15 Time-Varying Fields in Conductors (Magnetic Diffusion)

15.1 INTRODUCTION

During our study of the Maxwell's equations in Chapter 12, we found that these equations could be combined together to produce an operating equation in terms of a single field vector (i.e. the vector Helmholtz equation), which could be further reduced to two limiting equations. In the present chapter, we shall consider one of those two limiting cases, i.e. in the conducting media and at relatively low frequencies, the displacement current term can be neglected, so that the operating equation of our interest is

$$\nabla^2\mathbf{H} - j\omega\mu\sigma\mathbf{H} = 0 \tag{12.39a}$$

$$\nabla^2\mathbf{E} - j\omega\mu\sigma\mathbf{E} = 0 \tag{12.39b}$$

(assuming sinusoidal time-harmonic variations of the field vectors—though we shall consider some non-sinusoidal transient problems as well). These problems fall under the general category of eddy currents and skin effect types. In most of these problems, we can use the above equations as the starting point, though for the initial few problems, we shall derive the specific operating equation starting from the Maxwell's equations for the sake of clarity and a proper insight into the physics of the problems. In general, we shall evaluate the current distributions as well as the magnetic field distributions in the conducting bodies, followed by the energy dissipations or the losses and the energy transfers by both the Poynting vector as well as the current density distributions in the structures.

15.2 ALTERNATING CURRENT DISTRIBUTION IN A SEMI-INFINITE CONDUCTING BLOCK (SIMPLEST EXAMPLE OF THE VECTOR DIFFUSION-TYPE EQUATION)

Alternating current (ac) phenomena at a fixed frequency $f = \omega/2\pi$ are being investigated by the phasor method, in which we can write

$$\mathbf{J} = \mathrm{Re}\{\mathbf{J}\}\ \exp\ (j\omega t) \tag{15.1}$$

Here $\{\mathbf{J}\}$ stands for a set of three complex numbers, just as $\mathbf{J}$ is a set of three time-varying components. The fourth Maxwell's equation [i.e. Eq. (12.18)] then takes the form

$$\text{curl } \{\mathbf{H}\} = \nabla \times \{\mathbf{H}\} = \{\mathbf{J}\} + j\omega\{\mathbf{D}\} \tag{15.2}$$

[i.e. Eq. (12.22)].

Let us consider the relative magnitudes of the last two terms in a region filled with a material of conductivity σ and permittivity $\varepsilon/\varepsilon_0$.

Let $\mathbf{E}$ have a maximum value $\hat{E}$ in a certain direction, then $\mathbf{J}$ has the maximum value of $\sigma\hat{E}$ and $\mathbf{D}$ has the maximum value of $\varepsilon\hat{E}$ in the same direction.

$$\therefore \qquad \left|\frac{\text{Displacement current density}}{\text{Conduction current density}}\right| = \frac{\omega\varepsilon\hat{E}}{\sigma\hat{E}} = \omega\varepsilon\rho \tag{15.3}$$

where $\rho = 1/\sigma$.

For copper, $\rho = 1.7 \times 10^{-8}$ Ω-m and $\varepsilon = \varepsilon_0 = 8.854 \times 10^{-12}$.

The above ratio $\to 1$, when $f \to 10^{18}$ Hz (X-rays).

$\therefore$ For communication frequencies $< 10^6$ Hz, it is negligible in copper.

$\therefore$ For such problems, the displacement current can reasonably be neglected.

We now consider a semi-infinite block of resistivity ρ and permeability μ/μ_0 with one plane face taken as $x = 0$ (Figure 15.1). An alternating magnetizing force $\mathbf{H}_0$ is applied at this

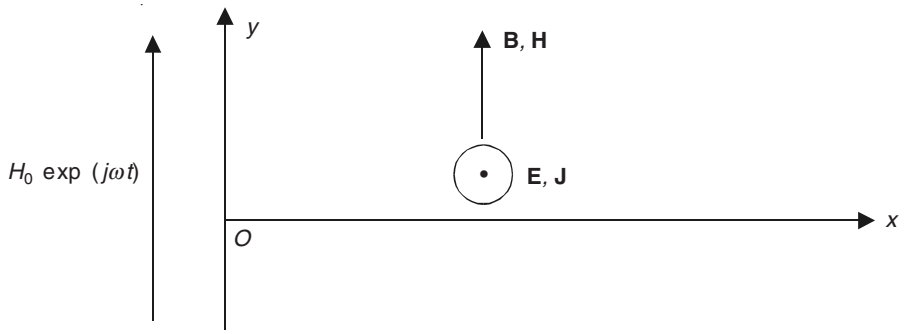


Figure 15.1 Semi-infinite conducting block subjected to an alternating magnetic field.

surface in the direction Oy, setting up $\mathbf{B}$ in the same direction in the material. The variations in $\mathbf{B}$ set up the electric fields and the consequent induced currents in the perpendicular Oz direction (i.e. normal to the plane of the paper). All the quantities will be phasors, so the notation $\{\mathbf{J}\}$ is not needed. For simplicity we write $\mathbf{B}$, $\mathbf{H}$, $\mathbf{J}$, $\mathbf{E}$ which are all assumed to be complex numbers, with the real quantity $\mathbf{H}_0$ acting as the reference phasor.

The relevant equations are:

$$\text{div } \mathbf{B} = \nabla \cdot \mathbf{B} = 0 \tag{15.4a}$$

$$\text{curl } \mathbf{E} = \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{15.4b}$$

$$\text{curl } \mathbf{H} = \nabla \times \mathbf{H} = \mathbf{J} \tag{15.4c}$$

$$\mathbf{E} = \rho\mathbf{J} \tag{15.4d}$$

$$\mathbf{B} = \mu\mathbf{H} \tag{15.4e}$$

with $\mathbf{B} = \mathbf{i}_x 0 + \mathbf{i}_y B_y + \mathbf{i}_z 0$ where $\mathbf{B}$ varies with x only, the zero divergence of $\mathbf{B}$ is automatically satisfied. Inserting the components into the other equations, and treating all the quantities as phasors so that the $\partial/\partial t$ operator becomes $j\omega$,

$$\frac{dE_z}{dx} = j\omega B_y \tag{15.5a}$$

$$\frac{dH_y}{dx} = J_z \tag{15.5b}$$

$$E_z = \rho J_z \tag{15.5c}$$

$$B_y = \mu H_y \tag{15.5d}$$

$$\therefore \qquad j\omega B_y = \frac{dE_z}{dx} = \rho\left(\frac{dJ_z}{dx}\right) = \rho\left(\frac{d^2 H_y}{dx^2}\right) = \left(\frac{\rho}{\mu}\right)\left(\frac{d^2 B_y}{dx^2}\right)$$

or

$$\frac{d^2 B_y}{dx^2} - j\left(\frac{\omega\mu}{\rho}\right) B_y = 0 \tag{15.6}$$

The other quantities H_y, E_z, J_z all obey equations of the same form. The dimensions of Eq. (15.6) show that the quantity $(\omega\mu/\rho)$ must have the dimensions of (1/length^2); we thus write

$$\left(\frac{\rho}{\omega\mu}\right)^{1/2} = d \tag{15.7}$$

and obtain

$$\left(\frac{d^2 B_y}{dx^2}\right) - \left(\frac{j}{d^2}\right) B_y = 0 \tag{15.8}$$

The general solution of this equation is

$$B_y = B_1 \exp\left(\frac{x\sqrt{j}}{d}\right) + B_2 \exp\left(\frac{-x\sqrt{j}}{d}\right) \tag{15.9}$$

where B_1 and B_2 are unknown constants of integration to be determined.

Boundary conditions to evaluate the unknowns B_1 and B_2. For a semi-infinite block, B_1 must be zero to keep the value of B_y finite as $x \to \infty$; also at $x = 0$, $B_y = B_0 = \mu H_0 = B_2$.

Thus

$$B_y = B_0 \exp\left(\frac{-x\sqrt{j}}{d}\right) \tag{15.10}$$

But

$$\sqrt{j} = \frac{1 + j}{\sqrt{2}}$$

So

$$B_y = B_0 \exp\left(\frac{-x\,(1 + j)}{d\sqrt{2}}\right)$$

or

$$B_y = B_0 \exp\left(\frac{-x}{d\sqrt{2}}\right)\left[\cos\left(\frac{x}{d\sqrt{2}}\right) - j\sin\left(\frac{x}{d\sqrt{2}}\right)\right] \tag{15.11}$$

In fact, if we include the time-variation term explicitly in the above expression,

$$B_y = B_0 \exp\left[\frac{-x(1+j)}{d\sqrt{2}}\right]\exp(j\omega t)$$

or

$$B_y = B_0 \exp\left(\frac{-x}{d\sqrt{2}}\right)\left[\cos\left(\frac{x}{d\sqrt{2}} - \omega t\right) - j\sin\left(\frac{x}{d\sqrt{2}} - \omega t\right)\right] \tag{15.12}$$

The first term $B_0 \exp\{-x/(d\sqrt{2})\}$ represents the amplitude of B_y as we penetrate into the block normal to the $x = 0$ surface, and the term in the bracket indicates the phase of B_y at each point x and at the instant of time t. The above expression shows that the vector **B** is a wave penetrating into the block moving forward in the x-direction, but as the wave penetrates further in, the amplitude of the wave is continually getting attenuated (Figure 15.2). Such a phenomenon is

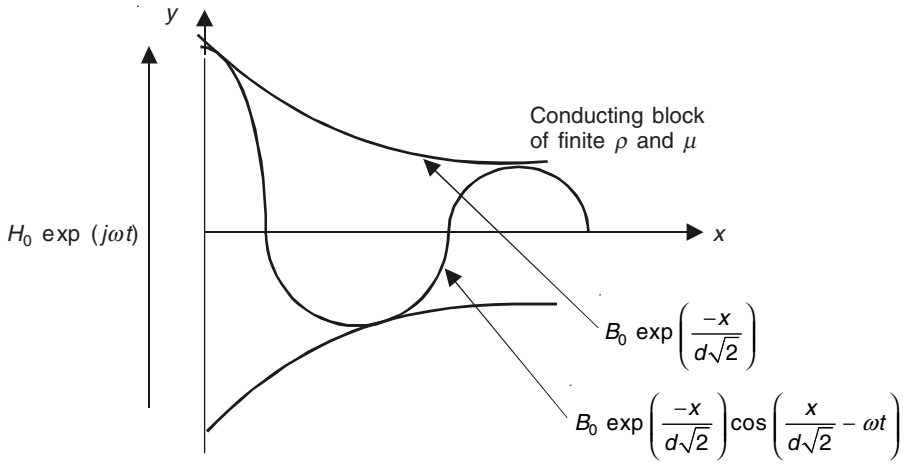


Figure 15.2 B_y wave diffusing into the conducting block.

called 'diffusion'. It will be seen that the magnetic field vector **B** (as all other field vectors) have the maximum value at the surface $x = 0$, and the magnitude decreases exponentially as x increases, the damping term (factor) being $\exp\{-x/(d\sqrt{2})\}$.

At a depth $x = d\sqrt{2}$, known as the 'depth of penetration', $B_y = (1/e)B_0$, i.e. B is $(1/e) = 0.368$ or 36.8% of the surface value of B (i.e. B_0 or B_S). The flux is thus mainly confined to the surface layers of the material. Next we calculate the total flux in the material, i.e. the total flux per unit length measured perpendicular to the plane of the diagram (= Φ) is given by

$$\Phi = \int_{x=0}^{x\to\infty} B_0 \exp\left(\frac{-x\sqrt{j}}{d}\right)dx = \frac{B_0 d}{\sqrt{j}}\left[-\exp\left(\frac{-x\sqrt{j}}{d}\right)\right]\Bigg|_0^\infty$$

$$= -\left(\frac{B_0 d\sqrt{2}}{1+j}\right)\left(\frac{1}{\infty}-1\right) = \frac{B_0 d(1-j)}{\sqrt{2}} \tag{15.13}$$

The term $(1 - j)/\sqrt{2}$ has the modulus unity, and shows that the total flux Φ lags the surface flux density by 45°. The multiplying term of B_0 in the above equation shows that the magnitude of the total flux Φ is the same as if the flux density at the surface (= B_0) had the constant value over the depth d. Thus d gives a measure of the effectiveness of the material as a carrier of the magnetic flux (or of current). On this basis, Professor Carter in his book, calls 'd' as the depth of penetration. But the more accepted definition of the depth of penetration is now $d\sqrt{2}$ which we also have accepted in our present discussion.

In copper, at 50 Hz, $\rho = 1.7 \times 10^{-8}$ Ω-m, $d\sqrt{2}$ = 0.93 cm. In iron, at the same frequency, taking $\rho = 10^{-7}$ Ω-m and $\mu = 10^3\mu_0$, $d\sqrt{2}$ = 0.071 cm. Thus there is no point in making the diameter of the wires, or the thickness of the transformer laminations, greater than about twice the depth of penetration at the operating frequency. This tendency for the alternating flux or current to crowd towards the surface of the material is called the 'skin effect'.

The other vectors may be derived from Eq. (15.10) via Eq. (15.5). Thus,

$$J_z = \frac{dH_y}{dx} = \left(\frac{1}{\mu}\right)\left(\frac{dB_y}{dx}\right) = -\left(\frac{B_0\sqrt{j}}{\mu d}\right)\exp\left(\frac{-x\sqrt{j}}{d}\right)$$

and

$$E_z = \rho J_z = -\left(B_0\omega d\sqrt{j}\right)\exp\left(\frac{-x\sqrt{j}}{d}\right) \tag{15.14}$$

since $\omega\mu/\rho = 1/d^2$ and $\omega d = \rho/(\mu d)$.

Just as we have plotted B_y as a function of x in the metal block (Figure 15.2), we shall now plot J_z (in the metal block) in the complex plane, to show the variations of the current density vector J_z, both in magnitude and phase (Figure 15.3), i.e.

$$J_z = -\left(\frac{B_0\sqrt{j}}{\mu d}\right)\exp\left(\frac{-x\sqrt{j}}{d}\right) = J_0\exp\left(\frac{-x}{d\sqrt{2}}\right)\exp\left(\frac{-jx}{d\sqrt{2}}\right)$$

$$= J_0\exp\left(\frac{-x}{d\sqrt{2}}\right)\angle -\theta$$

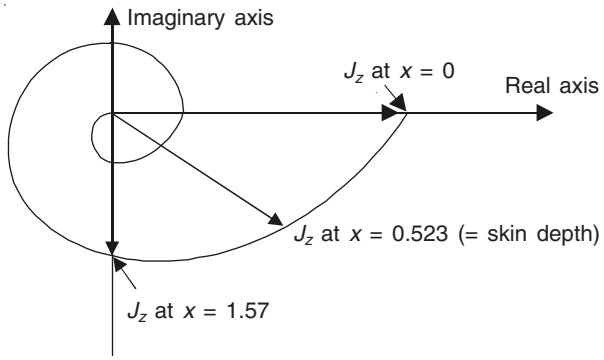


Figure 15.3 Complex plane diagram for J_z at various distances x from the surface of the semi-infinite metal block.

where

J_0 = the surface current density

$\theta = x/(d\sqrt{2})$.

In a circular wire of diameter $2a >> (d\sqrt{2})$, the current-carrying skin can be treated as a plane. (This, of course, is an approximation, which is justifiable.) Its width is $2\pi a$. Therefore, by analogy with Eq. (15.13), (Figure 15.4), the total current I is given by

$$I = 2\pi a J_0 d\left(\frac{1-j}{\sqrt{2}}\right) \tag{15.15}$$

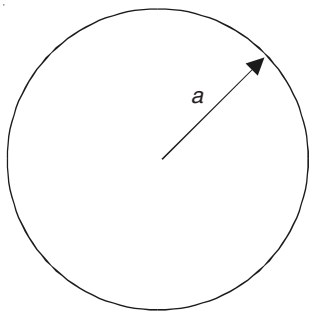


Figure 15.4 A metal wire of circular cross-section of radius $a >> d\sqrt{2}$.

where J_0 is the surface current density $= -\left(\frac{B_0\sqrt{j}}{\mu d}\right)$

or

$$I = 2\pi a d J_0\left(\frac{1-j}{\sqrt{2}}\right)$$

$\therefore$ The voltage drop V per unit length of the conductor is

$$V = \rho J_0 = \frac{\rho I}{2\pi a d}\left(\frac{1+j}{\sqrt{2}}\right) \tag{15.16}$$

We may thus write

$$V = (R + j\omega L)I, \qquad \text{where } R = \omega L = \frac{\rho}{2\sqrt{2}\pi a d} \tag{15.17}$$

The skin effect thus increases the resistance of the wire, and in addition causes it to have inductance. The resistance R_0 per unit length, at very low frequencies (and for dc at $f = 0$) is $\rho/(\pi a^2)$.

$\therefore$

$$\frac{R_{ac}}{R_0} = \frac{\omega L}{R_0} = \frac{a}{2\sqrt{2}d} \tag{15.18}$$

A more accurate analysis in cylindrical coordinates (which we shall see subsequently) gives this ratio as $[a/(2\sqrt{2}\,d) + 1/4]$. Thus, this simplified analysis gives a good approximation to the ratio R/R_0, when a/d exceeds about 3.

15.2.1 Power Loss Calculation (Using Poynting Vector)

We use the complex Poynting vector:

$$\mathbf{S}' = \mathbf{E}_c \times \mathbf{H}_c^*$$

At the surface $x = 0$, $H_y = H_0$, and $E_z = -E_0 = -B_0 \omega d \sqrt{j}$ [from Eq. (15.14)]. See Figure 15.5. Thus $(\mathbf{E} \times \mathbf{H})_0$ is normally inwards, with no corresponding outward vector at $x \to \infty$. The power

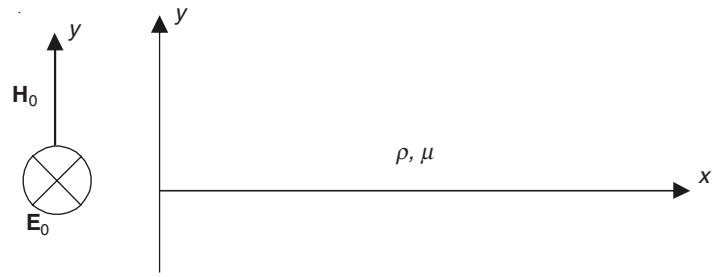


Figure 15.5 Surface field values of the semi-infinite block.

denoted by $(\mathbf{E} \times \mathbf{H}_0)$ is the ohmic loss associated with the eddy currents. In this case

$$\mathbf{H}^* = \mathbf{H} = \mathbf{H}_0 = \mathbf{i}_y H_0$$

and

$$\mathbf{E} = \mathbf{E}_0 = \mathbf{i}_z \mu H_0 \omega d \sqrt{j} = \mathbf{i}_z \mu H_0 \omega d \left(\frac{1+j}{\sqrt{2}} \right)$$

$\therefore$ The complex Poynting vector $\mathbf{S}'$ is

$$\mathbf{S}' = \mathbf{E} \times \mathbf{H}^* = \mathbf{i}_x H_0^2 \omega \mu d \left(\frac{1+j}{\sqrt{2}} \right)$$

and the average value is

$$\bar{S} = \frac{1}{2} \operatorname{Re}(\mathbf{S}') = H_0^2 \frac{\omega \mu d}{2\sqrt{2}} \text{ watts/metre}^2 \tag{15.19}$$

Substituting for $d = \sqrt{\rho/(\omega\mu)}$,

$$\bar{S} = \left(\frac{H_0^2}{2\sqrt{2}} \right) \sqrt{\omega \mu \rho} \text{ watts/metre}^2 \tag{15.20}$$

To evaluate this without recourse to the Poynting vector, we consider the induced current density $\mathbf{J}$,

$$\mathbf{J} = -\mathbf{i}_z \left(H_0 \frac{\sqrt{j}}{d} \right) \exp\left(\frac{-x\sqrt{j}}{d} \right)$$

which has the magnitude

$$|\mathbf{J}| = \left(\frac{H_0}{d} \right) \exp\left(\frac{-x}{d\sqrt{2}} \right)$$

The loss is then

$$= \int_0^\infty \left(\frac{1}{2} \right) \rho |\mathbf{J}|^2 \, dx \tag{15.21}$$

(where the factor (1/2) comes in because the peak values are assumed).

$$= \frac{1}{2}\rho\left(\frac{H_0^2}{d^2}\right)\int_0^\infty \exp\left(\frac{-x\sqrt{2}}{d}\right)dx$$

$$= \frac{1}{2}\rho\left(\frac{H_0^2}{d^2}\right)\left(\frac{d}{\sqrt{2}}\right)\left[-\exp\left(\frac{-x\sqrt{2}}{d}\right)\right]\Bigg|_0^\infty$$

$$= \frac{\rho H_0^2}{2\sqrt{2}d} = \left(\frac{H_0^2}{2\sqrt{2}}\right)(\omega\mu\rho)^{1/2}$$

which is same as the previous Eq. (15.20). The complex Poynting vector thus gives a quicker solution, partly because one needs only use the values of the field quantities at the surface, not all through the material.

15.3 PENETRATION OF FLUX AND CURRENT INTO A PLATE OF FINITE THICKNESS (INDUCTION HEATING)

From the analysis of a semi-infinite block, it is a direct extrapolation to solve for a plate of finite thickness with two parallel plane faces (Figure 15.6). The origin of the system is taken at the

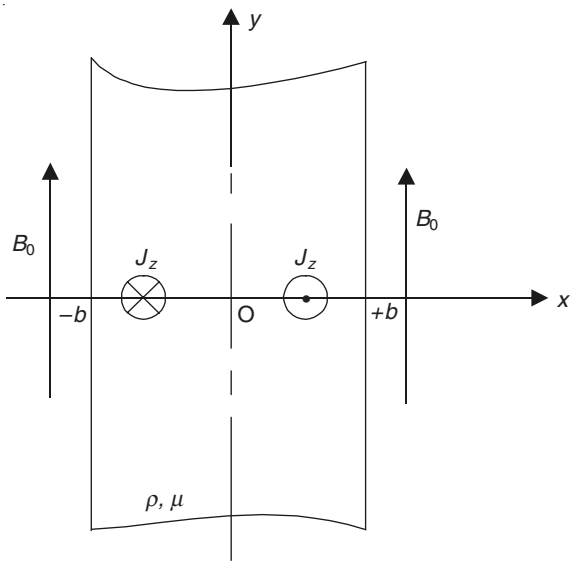


Figure 15.6 Plane faced metal block.

centre of the plate, so that the plane faces are given by $x = \pm b$. The same field is applied on both the faces, i.e. $B = B_0$ at $x = \pm b$.

The initial steps of the analysis are similar to those of the semi-infinite block, and the general solution comes out of the form

$$B_y = B = B_1 \exp\left(\frac{x\sqrt{j}}{d}\right) + B_2 \exp\left(\frac{-x\sqrt{j}}{d}\right)$$

corresponding to Eq. (15.9). The constants B_1 and B_2 have to be evaluated using the conditions that $B_y = B_0$ at $x = \pm b$. This gives

$$B_y = B_0 \left[\frac{\exp\left(\frac{x\sqrt{j}}{d}\right) + \exp\left(\frac{-x\sqrt{j}}{d}\right)}{\exp\left(\frac{b\sqrt{j}}{d}\right) + \exp\left(\frac{-b\sqrt{j}}{d}\right)} \right]$$

which can be expressed as

$$B_y = B_0 \left[\frac{\cosh\left(\frac{x\sqrt{j}}{d}\right)}{\cosh\left(\frac{b\sqrt{j}}{d}\right)} \right]$$

$$= B_0 \frac{\cosh\left[\left(\frac{x}{d}\right)\left(\frac{1+j}{\sqrt{2}}\right)\right]}{\cosh\left[\left(\frac{b}{d}\right)\left(\frac{1+j}{\sqrt{2}}\right)\right]} \tag{15.22}$$

If such a plate is to be used as a lamination in a transformer core, then the deciding factor would be the total amount of magnetic flux, which would be carried by the unit depth of the lamination. This would be the basis for deciding the minimum thickness of each lamination so that its central layers are optimally used.

The value of the total flux per unit length (= Φ) is given by

$$\Phi = \int_{x=-b}^{x=+b} B_y \, dx$$

$$= \frac{B_0}{\cosh\left[\left(\frac{b}{d}\right)\left(\frac{1+j}{\sqrt{2}}\right)\right]} \int_{x=-b}^{x=+b} \cosh\left[\left(\frac{x}{d}\right)\left(\frac{1+j}{\sqrt{2}}\right)\right] dx$$

$$= \frac{2B_0 d\left(\frac{\sqrt{2}}{1+j}\right) \sinh\left[\left(\frac{b}{d}\right)\left(\frac{1+j}{\sqrt{2}}\right)\right]}{\cosh\left[\left(\frac{b}{d}\right)\left(\frac{1+j}{\sqrt{2}}\right)\right]}$$

$$= 2B_0 d\left(\frac{1-j}{\sqrt{2}}\right) \times \frac{\sinh\alpha\cos\alpha + j\cosh\alpha\sin\alpha}{\cosh\alpha\cos\alpha + j\sinh\alpha\sin\alpha} \tag{13.23}$$

where

$$a = \frac{b}{d\sqrt{2}} \tag{15.23}$$

The magnitude of this flux is $|\Phi|$, which is

$$|\Phi| = 2B_0 d \times \sqrt{\frac{\sinh^2\alpha\cos^2\alpha + \cosh^2\alpha\sin^2\alpha}{\cosh^2\alpha\cos^2\alpha + \sinh^2\alpha\sin^2\alpha}}$$

$$= 2B_0 d \times \sqrt{\frac{\cosh 2\alpha - \cos 2\alpha}{\cosh 2\alpha + \cos 2\alpha}}$$

If the flux penetration were complete, then the flux would be

$$2B_0 b = \Phi_0$$

$$\therefore \quad |\Phi| = \left(\frac{\Phi_0}{\alpha\sqrt{2}}\right) \times \sqrt{\frac{\cosh 2\alpha - \cos 2\alpha}{\cosh 2\alpha + \cos 2\alpha}} \tag{15.24}$$

The ratio $|\Phi/\Phi_0|$ is a measure of the reduction in the effective permeability of the lamination material due to the presence of the eddy currents in it. This is plotted as a function of (b/d) in

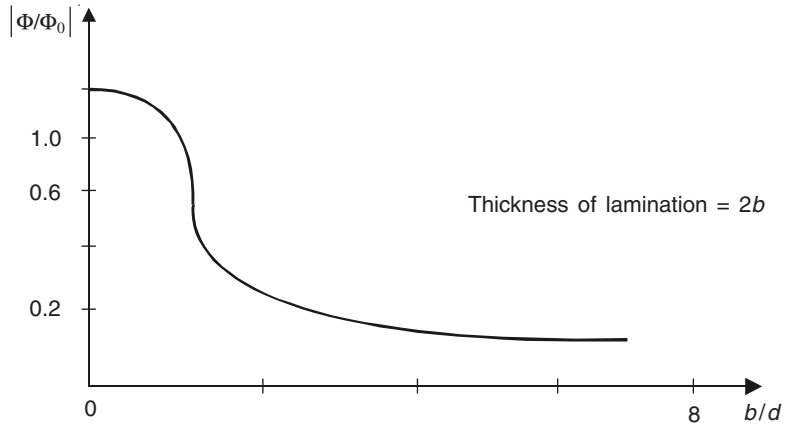


Figure 15.7 Penetration of Φ as a function of b.

Figure 15.7. The next parameter important from the designer's point of view is the eddy current loss in the laminations, and so we now calculate the eddy current loss. So we need to evaluate the current density vector **J**, i.e.

$$J_z = -\frac{dH_y}{dx} = -\left(\frac{1}{\mu_0\mu_r}\right)\left(\frac{dB_y}{dx}\right)$$

and using the value of B_y obtained in Eq. (15.22), we get

$$J_z = \left(\frac{B_0}{\mu_0\mu_r d}\right)\left(\frac{1+j}{\sqrt{2}}\right) \times \frac{\sinh\left[\left(\frac{x}{d}\right)\left(\frac{1+j}{\sqrt{2}}\right)\right]}{\cosh\left[\left(\frac{b}{d}\right)\left(\frac{1+j}{\sqrt{2}}\right)\right]}$$

which can be expressed as

$$J_z = -\left(\frac{B_0}{\mu_0 \mu_r d}\right)\left(\frac{1+j}{\sqrt{2}}\right) \times \frac{\sinh\left(\frac{x}{d\sqrt{2}}\right)\cos\left(\frac{x}{d\sqrt{2}}\right) + j\cosh\left(\frac{x}{d\sqrt{2}}\right)\sin\left(\frac{x}{d\sqrt{2}}\right)}{\cosh\alpha\cos\alpha + j\sinh\alpha\sin\alpha}$$

and hence the magnitude of J_z is

$$J_z = \left(\frac{B_0}{\mu_0 \mu_r d}\right) \times \sqrt{\frac{\sinh^2\left(\frac{x}{d\sqrt{2}}\right)\cos^2\left(\frac{x}{d\sqrt{2}}\right) + \cosh^2\left(\frac{x}{d\sqrt{2}}\right)\sin^2\left(\frac{x}{d\sqrt{2}}\right)}{\cosh^2\alpha\cos^2\alpha + \sinh^2\alpha\sin^2\alpha}}$$

$$= \left(\frac{B_0}{\mu_0 \mu_r d}\right) \times \left[\frac{\cosh\left(\frac{x\sqrt{2}}{d}\right) - \cos\left(\frac{x\sqrt{2}}{d}\right)}{\cosh 2\alpha + \cos 2\alpha}\right]$$

∴ The eddy current loss (= W_e) is

$$W_e = \int_{-b}^{+b} \rho |J_z|^2 \, dx$$

$$= \left(\frac{\rho B_0^2 \sqrt{2}}{\mu_0 \mu_r d}\right)\left(\frac{\sinh 2\alpha - \sin 2\alpha}{\cosh 2\alpha + \cos 2\alpha}\right) \tag{15.25}$$

It will be seen that the second term is the square of the term for the flux-linkage as obtained in Eq. (15.24).

A second application of the results of this problem is in the design and use of induction furnaces. The term, [(sinh 2α − sin 2α)/(cosh 2α + cos 2α)] is the important parameter for the users of induction furnaces. The above quantity is a function of α or $b/(d\sqrt{2})$, say $F(\alpha)$ or $F\{b/(d\sqrt{2})\}$. If this function is plotted for a range of $b/(d\sqrt{2})$ or α, it will be seen that for maximum heating, the sheets should have a thickness $2b \simeq 2.25(d\sqrt{2})$, i.e. b should be slightly greater than the skin-depth ($d\sqrt{2}$) ($=\delta$). If the sheets are thicker, the eddy currents will not penetrate far enough; and if the sheets are thinner, the eddy currents will be smaller.

15.4 ALTERNATING CURRENT FLOW IN A FLAT SHEET

We now look at the problem of an alternating current flowing in a sheet of finite thickness ($2b$) (Figure 15.8). In this case, the current flows in the y-direction and hence the current density vector will have only the y-component, and the magnetic flux density **B** will have only the z-component. These field vectors will vary only with x, i.e. $\partial/\partial y = 0$ and $\partial/\partial z = 0$.

∴ curl **H** = **J** equation will reduce to

$$-\frac{\partial H_z}{\partial x} = J_y \tag{15.26}$$

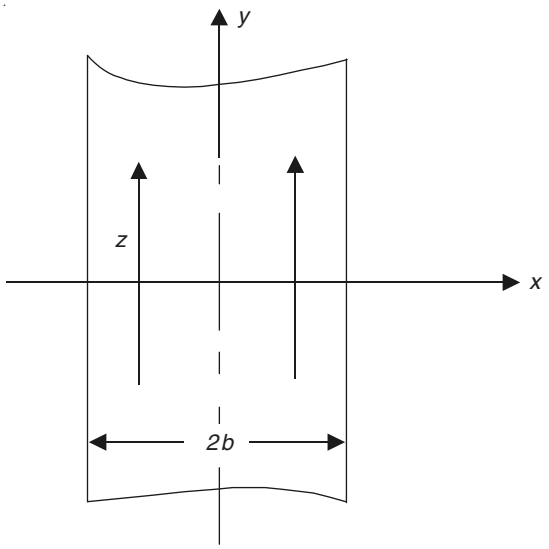


Figure 15.8 Current flow in a flat sheet.

and curl $\mathbf{E} = -\dfrac{\partial \mathbf{B}}{\partial t}$ reduces to

$$\frac{\partial E_y}{\partial x} = -\frac{\partial B_z}{\partial t} \tag{15.27}$$

Using the constituent relations

$$\mathbf{B} = \mu \mathbf{H} \text{ and } \mathbf{J} = \sigma \mathbf{E} \tag{15.28}$$

we get

$$J_y = -\left(\frac{1}{\mu_0}\right)\left(\frac{\partial B_z}{\partial x}\right) \quad \text{and} \quad \left(\frac{\partial B_z}{\partial t}\right) = \left(\frac{1}{\sigma}\right)\left(\frac{\partial J_y}{\partial x}\right)$$

$$\therefore \quad \frac{\partial J_y}{\partial t} = -\left(\frac{1}{\mu_0}\right)\left(\frac{\partial^2 B_z}{\partial x \partial t}\right) = +\left(\frac{1}{\mu_0 \sigma}\right)\left(\frac{\partial^2 J_y}{\partial x^2}\right) \tag{15.29}$$

Since the current through the block is alternating, we write it as

$$J_y = \hat{J} \exp\left(j\omega t\right) \tag{15.30}$$

$\therefore$ Equation (15.29) becomes

$$\frac{\partial^2 \hat{J}}{\partial x^2} = j\omega\mu_0\sigma\hat{J} \tag{15.31}$$

Since all the quantities are phasors, we simplify our writing by omitting the peak notation, and as before

$$\frac{1}{\omega\mu_0\sigma} = d^2 \tag{15.32}$$

the equation for the current density vector becomes

$$\frac{\partial^2 J_y}{\partial x^2} - \left(\frac{j}{d^2}\right) J_y = 0 \tag{15.33}$$

Its general solution is (i.e. the space variation of J),

$$J_y = J_1 \exp\left(\frac{x\sqrt{j}}{d}\right) + J_2 \exp\left(\frac{-x\sqrt{j}}{d}\right) \tag{15.34}$$

where J_1 and J_2 are unknown constants to be determined from the boundary conditions which are:

(1) at $x = 0$, $J_y = J_0$ (say); and
(2) $x = 0$ is the axis of symmetry.

Hence, from these conditions, the two equations are

$$J_0 = J_1 + J_2 \qquad \text{and} \qquad J_1 = J_2$$

$\therefore$ The solution becomes

$$J_y = \left(\frac{J_0}{2}\right)\left[\exp\left\{\frac{(1+j)x}{d\sqrt{2}}\right\} + \exp\left\{\frac{-(1+j)x}{d\sqrt{2}}\right\}\right]$$
$$= J_0 \cosh\left[\frac{(1+j)x}{d\sqrt{2}}\right] \tag{15.35}$$

On the surface $x = \pm b$,

$$J_y = J_S = J_0 \cosh\left[\frac{(1+j)b}{d\sqrt{2}}\right]$$

$\therefore$ Expressing the current density vector in terms of the surface value,

$$J_y = J_S\left[\frac{\cosh\left\{\frac{(1+j)x}{d\sqrt{2}}\right\}}{\cosh\left\{\frac{(1+j)b}{d\sqrt{2}}\right\}}\right] \tag{15.36}$$

The total current per unit width of the sheet is

$$I = \int_{-b}^{+b} J_y\, dx = \left(\frac{2\sqrt{2}J_S d}{1+j}\right)\tanh\left(\frac{(1+j)b}{d\sqrt{2}}\right) \tag{15.37}$$

If $b/(d\sqrt{2})$ is large, then $\tanh\,[(1+j)b/(d\sqrt{2})] \to 1$. This is true within 1% if $b/(d\sqrt{2}) > 2.64$. Thus in a sheet of this thickness,

$$I = \frac{2\sqrt{2}J_S d}{1+j} \tag{15.38}$$

We now calculate the impedance of the sheet, which would also give us the loss in an alternative way. The voltage applied to the sheet can be obtained from the surface electric field.

The voltage per unit length in the y-direction is

$$V = \frac{J_S}{\sigma} \tag{15.39}$$

∴ The complex impedance per unit area of y- and z-directions is

$$Z = \frac{V}{I} = \left(\frac{1+j}{2\sqrt{2}d\sigma}\right) \coth\left[\frac{(1+j)b}{d\sqrt{2}}\right] \tag{15.40}$$

Separating the real and the imaginary parts,

$$Z = R + jX = \left(\frac{1+j}{2\sqrt{2}d\sigma}\right) \times \frac{\cosh\alpha\cos\alpha + j\sinh\alpha\sin\alpha}{\sinh\alpha\cos\alpha + j\cosh\alpha\sin\alpha}$$

where

$$\alpha = \frac{b}{d\sqrt{2}} \tag{15.41}$$

After further manipulations,

$$R = \left(\frac{1}{2\sqrt{2}d\sigma}\right)\left(\frac{\sinh 2\alpha + \sin 2\alpha}{\cosh 2\alpha - \cos 2\alpha}\right)$$

$$X = \left(\frac{1}{2\sqrt{2}d\sigma}\right)\left(\frac{\sinh 2\alpha - \sin 2\alpha}{\cosh 2\alpha - \cos 2\alpha}\right) \tag{15.42}$$

This reactance is due to the redistribution of the current inside the sheet, and hence is due to the 'internal inductance' of the sheet. The resistance in Eq. (15.42) is due to the alternating current. The direct current resistance of the sheet is

$$R_{dc} = \frac{1}{2b\sigma} \tag{15.43}$$

∴ The ratio of ac to dc resistance is

$$\frac{R_{ac}}{R_{dc}} = \frac{b}{d\sqrt{2}} \frac{\sinh 2\alpha + \sin 2\alpha}{\cosh 2\alpha - \cos 2\alpha} \tag{15.44}$$

When this ratio is plotted as a function of $b/(d\sqrt{2})$, then for large values of $\alpha = b/(d\sqrt{2})$, it tends to $b/(d\sqrt{2})$ (Figure 15.9), because the current tends to concentrate in the surface layers of $b/(d\sqrt{2})$. At low values of $b/(d\sqrt{2})$, the expression can be expanded in the ascending powers of α, i.e.

$$\frac{R_{ac}}{R_{dc}} = 1 + \left(\frac{4}{45}\right)\alpha^4 \tag{15.45}$$

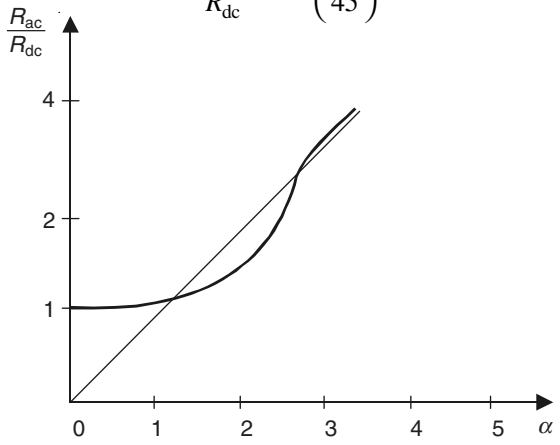


Figure 15.9 AC resistance of a flat sheet.

When we consider thin sheets, it is seen that the above expression still holds, i.e. the ac loss for thin sheets can be evaluated without considering the effects of the magnetic field of the eddy currents. If the effects of the field of eddy currents are negligible, then the behaviour is said to be 'resistance-limited'. For thick sheets, in which the current is confined in the surface layers, the behaviour is said to be 'inductance-limited'.

15.5 SKIN EFFECT (OR EDDY CURRENTS) IN AN ISOLATED CIRCULAR CONDUCTOR

We consider a long conductor, of circular cross-section, carrying an alternating current I, the current density at any radius r being $\mathbf{J}$, the direction of the current being normal to the plane of the paper into it (Figure 15.10).

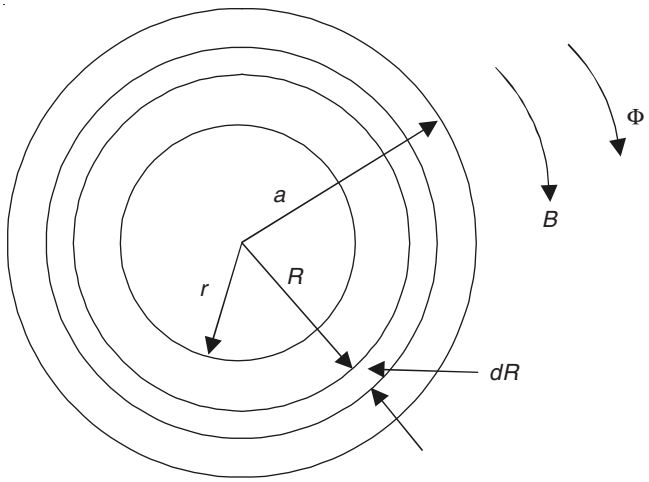


Figure 15.10 A circular conductor of radius *a*, carrying an alternating current.

The total current = I; the current density = $\mathbf{J}$; the applied voltage = V; the induced voltage = e; the frequency of the current = f; or angular frequency = $\omega = 2\pi f$; the external flux linked per unit length = Φ.

The current density is a function of the radius only. The alternating current gives rise to the magnetic flux; and from the Maxwell's equation,

$$\oint_C H_r \, dl = \text{current enclosed by the contour,}$$

$$2\pi RH(R) = \int_{r=0}^{r=R} 2\pi r dr J(r) \tag{15.46}$$

This equation is true at any instant of time, since all are functions of time.

Differentiating with respect to R,

$$\frac{\partial}{\partial R}(RH) = RJ \tag{15.47}$$

[At the centre ($r \to 0$), $H \times 2\pi r = \pi r^2 J_0$. $\therefore \quad H = (J_0/2)r \to$ as $r \to 0$.]

Let Φ = the flux linking the conductor externally per unit length. Then the total flux linking with the core of radius R is given by

$$\text{External flux } \Phi + \int_R^{r=a} \mu_0 \mu_r H \, 1 \cdot dr$$

$\therefore$ Induced voltage per unit length at radius r is

$$\equiv e = -\frac{\partial}{\partial t}\left[\Phi + \int_R^a \mu_0 \mu_r H dr\right] \tag{15.48}$$

This creates a certain current density, in addition to the applied voltage. We know that V = applied voltage per unit length.

$\therefore$ $\underbrace{V + e}_{\text{Total voltage}} = \underbrace{\rho J}_{\text{Gradient}}$ (i.e. resistivity $\times$ current density)

$$\therefore \quad e = \rho J - V$$

Let us assume that Φ is finite.

$\therefore$
$$-\frac{\partial}{\partial t}\left[\Phi + \int_R^a \mu_0 \mu_r H dr\right] = \rho J - V = e \tag{15.49}$$

Differentiating with respect to R,

$$-\frac{\partial}{\partial t}\left[-\mu_0 \mu_r H(R)\right] = \rho\left(\frac{dJ}{dR}\right) - 0$$

or

$$\mu_0 \mu_r \left(\frac{\partial H}{\partial t}\right) = \rho\left(\frac{dJ}{dR}\right) \tag{15.50}$$

Since the current flowing through the conductor is assumed to be sinusoidal, $\partial/\partial t = j\omega$ and Eq. (15.50) becomes

$$\left(\frac{j\omega\mu_0\mu_r}{\rho}\right) H = \frac{dJ}{dR}$$

or

$$j\lambda^2 H = \frac{dJ}{dR}$$

where

$$\lambda^2 = \frac{\omega\mu_0\mu_r}{\rho} = \frac{1}{d^2} \tag{15.51}$$

$\therefore$ From Eqs. (15.47) and (15.51),

$$j\lambda^2 H = \frac{d}{dR}\left[\left(\frac{1}{R}\right)\frac{\partial}{\partial R}(RH)\right] = \frac{d}{dR}\left(\frac{dH}{dR} + \frac{H}{R}\right)$$

or

$$R^2\left(\frac{d^2H}{dR^2}\right) + R\left(\frac{dH}{dR}\right) - \left(j\lambda^2 R^2 + 1\right)H = 0$$

Compare this equation with the Bessel's equation of the order n, which is

$$x^2\left(\frac{d^2y}{dx^2}\right)+x\left(\frac{dy}{dx}\right)+\left(x^2-n^2\right)y=0 \tag{15.52}$$

so that the equation for H can be rewritten as

$$R^2\left(\frac{d^2H}{dR^2}\right)+R\left(\frac{dH}{dR}\right)+\left(j^3\lambda^2R^2-1\right)H=0 \tag{15.53}$$

$\therefore$ The solution of this equation is

$$H = AJ_1(j^{3/2}\lambda R) + BY_1(j^{3/2}\lambda R) \tag{15.54}$$

where $J_1(j^{3/2}\lambda R)$ and $Y_1(j^{3/2}\lambda R)$ are Bessel's functions (of first order) of first and second kind respectively. It should be noted that the arguments of these Bessel's functions are complex and hence they can be expressed as Kelvin functions, i.e. Ber and Bei functions. However we shall use them as they are and convert them later to polar form. (Note that A and B are unknown constants of integration, and can be complex.)

From Eq. (15.47), the current density J can now be expressed as

$$J = \left(\frac{1}{R}\right)\left[\frac{d}{dR}(RH)\right]$$

$$= \left(\frac{1}{R}\right)\left[\frac{d}{dR}\left\{RAJ_1\left(j^{3/2}\lambda R\right)+RBY_1\left(j^{3/2}\lambda R\right)\right\}\right]$$

$$= \left(\frac{1}{R}\right)\times\frac{d}{d\left(j^{3/2}\lambda R\right)}\left[\left(j^{3/2}\lambda R\right)AJ_1\left(j^{3/2}\lambda R\right)+\left(j^{3/2}\lambda R\right)BY_1\left(j^{3/2}\lambda R\right)\right]$$

$$= \left(\frac{1}{R}\right)\left[\left(j^{3/2}\lambda R\right)AJ_0\left(j^{3/2}\lambda R\right)+\left(j^{3/2}\lambda R\right)BY_0\left(j^{3/2}\lambda R\right)\right]$$

$\therefore$ $$J = \left[\left(j^{3/2}\lambda\right)AJ_0\left(j^{3/2}\lambda R\right)+\left(j^{3/2}\lambda\right)BY_0\left(j^{3/2}\lambda R\right)\right]$$

Note that $j^{3/2}$ implies a phase shift.

The Bessel's function of the second kind $Y_n(x) \to \infty$ as $x \to 0$.

But $H \to 0$ as $R \to 0$. $\therefore$ $B = 0$.

$$\therefore \quad H = AJ_1(j^{3/2}\lambda R), \qquad J = (j^{3/2}\lambda)\,AJ_0(j^{3/2}\lambda R)$$

However in tubular conductors, the general solution for the finite value of B is valid.

Now, to evaluate A, if the total current is I, then at $R = a$, $\oint Hdr = I$; or

$$2\pi aH_a = I = 2\pi aAJ_1(j^{3/2}\lambda a)$$

$$\therefore \quad A = \frac{I}{2\pi aJ_1\left(j^{3/2}\lambda a\right)}$$

$$\therefore \qquad H = \left(\frac{I}{2\pi a}\right)\frac{J_1\left(j^{3/2}\lambda R\right)}{J_1\left(j^{3/2}\lambda a\right)}, \quad \text{and} \quad J = \frac{\left(j^{3/2}\lambda\right)I}{2\pi a}\,\frac{J_0\left(j^{3/2}\lambda R\right)}{J_1\left(j^{3/2}\lambda a\right)} \tag{15.55}$$

Expressing the Bessel's functions in polar form

$$H = \left(\frac{I}{2\pi a}\right)\left[\frac{M_1(\lambda R)\exp\{j\theta_1(\lambda R)\}}{M_1(\lambda a)\exp\{j\theta_1(\lambda a)\}}\right]$$

$$J = \left[\frac{\left(j^{3/2}\lambda\right)I}{2\pi a}\right]\left[\frac{M_0(\lambda R)\exp\{j\theta_0(\lambda R)\}}{M_1(\lambda a)\exp\{j\theta_1(\lambda a)\}}\right] \tag{15.56}$$

We now plot the amplitude $M_0(\lambda R)$ as a function of $(\lambda R) = x$ (Figure 15.11). From the curve, it will be seen that for sufficiently high λ (i.e. the reciprocal of skin-depth/$\sqrt{2}$), at the outer radii, the current density increases.

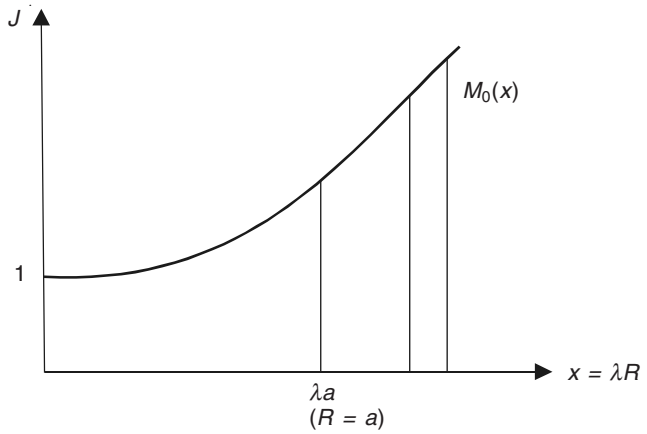


Figure 15.11 Current density amplitude as a function of radius R of the conductor.

$\therefore$ At higher frequencies (f), most of the current will be located in the outermost shells. The current density at the centre of the wire is minimum, and increases monotonically with radius. Furthermore, if '(λa)' is increased, the discrepancy of the current density at the centre and the outer radius increases and vice versa.

Next, we consider the phase of the current density vector. The terms affecting the phase are (Figure 15.12)

$$j^{3/2} \rightarrow \left(\frac{3\pi}{4}\right) + \theta_0(\lambda R) - \theta_1(\lambda a) = \text{phase of } J - \text{phase of } I \tag{15.57}$$

Note: For large values of $x(= \lambda R)$, $x > 45$

$\theta_0(x) = 40 \cdots x - 5.06/x - 22½°$; $\qquad \theta_1(x) = 40 \cdots x + 15.19/x + 67½°$.

When $\lambda \rightarrow 0$ ($\lambda = \omega\mu_0\mu_r/\rho$), i.e. at low frequencies, all the current densities (at different radii) are in phase. At sufficiently high frequencies, in the current density at any radius, there are different phase reversals.

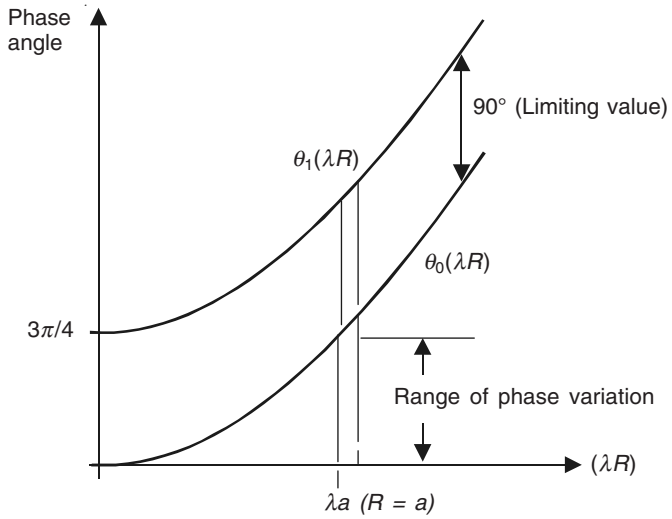


Figure 15.12 Phase variations in J as a function of R.

(*Note:* We have already seen that the skin current density leads the resultant total current by 45°—true for high frequencies as well.) Note further that at high frequencies, a (very very) small change in r is sufficient to cause the phase change by 45° between two shells.

15.5.1 Loss Calculation and the Effective Resistance in the Circular Conductor

We now calculate the effective resistances for various values of λ.

$$J = \left[\frac{(j^{3/2}\lambda)I}{2\pi a}\right]\left[\frac{M_0(\lambda R)\exp\{j\theta_0(\lambda R)\}}{M_1(\lambda a)\exp\{j\theta_1(\lambda a)\}}\right] \quad \text{at radius } R \tag{15.58}$$

$$\therefore \quad |J| = \left[\frac{\lambda|I|}{2\pi a}\right]\left[\frac{M_0(\lambda R)}{M_1(\lambda a)}\right]$$

where

$$|I| = \text{amplitude of the current} \tag{15.59}$$

$\therefore$ The total ohmic loss per unit length =

$$\underbrace{\left[\frac{\rho 1}{2\pi r dr}\right]}_{\substack{\uparrow \\ \text{resistance} \\ \text{per unit length}}}\underbrace{\left[\frac{(2\pi r dr)^2}{2}\right]}_{\substack{\uparrow \\ (\text{area})^2}}\underbrace{\left[\frac{\lambda^2|I|^2}{4\pi^2 a^2}\right]\left[\frac{M_0^2(\lambda R)}{M_1^2(\lambda a)}\right]}_{\substack{\uparrow \\ (\text{current density})^2}} \tag{15.60}$$

i.e. the loss is proportional to the (rms current)^2.

This is to be integrated within the limits 0 to a, thus, the loss

$$= \frac{\rho\lambda^2|I|^2}{(4\pi a^2)M_1^2(\lambda a)}\int_0^a rM_0^2(\lambda r)dr \tag{15.61}$$

We note that, when $\lambda = 0$, i.e. working under dc conditions, resistance per unit length

$$R_{dc} = \frac{\rho}{\pi a^2}$$

$$\therefore \quad \text{DC losses} = \left(\frac{\rho}{\pi a^2}\right)\left(\frac{|I|^2}{2}\right) \tag{15.62}$$

$\therefore$ Multiplying factor (for AC due to $\lambda \neq 0$)

$$= \left[\frac{\lambda^2}{2M_1^2(\lambda a)}\right]\int_0^a r M_0^2(\lambda r)dr \tag{15.63}$$

$$\left[\text{at } \lambda = 0,\ M_1^2(\lambda a) \to \left(\frac{\lambda a}{2}\right)^2 \qquad \text{and} \qquad M_0(\lambda r) \to 1\right]$$

We refer to Lommel's integral,

$$\int_0^a rJ_0(kr)J_0(lr)dr = \left(\frac{a}{k^2 - l^2}\right)[(kJ_0(la)J_1(ka) - lJ_0(ka)J_1(la)] \tag{15.64}$$

In this expression, let $k = j^{3/2}\lambda$, $l = (-j)^{-3/2}\lambda = j^{-3/2}\lambda$.

On substituting and expressing the integral in the polar form:

$$\int_0^a M_0(\lambda r)\exp\{-j\theta_0(\lambda r)\}\, M_0(\lambda r)\exp\{-j\theta_0(\lambda r)\}rdr = \left(\frac{a}{j^3\lambda^2 - j^{-3}\lambda^2}\right)\times$$

$$[(j^{3/2}\lambda)M_0(\lambda a)\exp\{-j\theta_0(\lambda a)\}\, M_1(\lambda a)\exp\{j\theta_1(\lambda a)\}$$

$$-(j^{-3/2}\lambda)M_0(\lambda a)\exp\{j\theta_0(\lambda a)\}M_1(\lambda a)\exp\{-j\theta_1(\lambda a)\}]$$

noting that $j^3 = -j$ and $j^{-3} = j$.

$$\int_0^a rM_0^2(\lambda r)dr$$

$$= -\left(\frac{a}{2j\lambda}\right)[\exp j\{(3\pi/4) + \theta_1(\lambda a) - \theta_0(\lambda a)\}] - \exp[-j\{(3\pi/4) + \theta_1(\lambda a) - \theta_0(\lambda a)\}]\times[M_0(\lambda a)M_1(\lambda a)]$$

noting that $j^{3/2} = \exp\{j(3\pi/4)\}$ and $j^{-3/2} = \exp\{-j(3\pi/4)\}$

$$= -\left(\frac{a}{\lambda}\right)[\sin\{(3\pi/4) + \theta_1(\lambda a) - \theta_0(\lambda a)\}][M_0(\lambda a)\, M_1(\lambda a)]$$

$$= -\left(\frac{a}{\lambda}\right)[\sin\{-(\pi/4) + \theta_1(\lambda a) - \theta_0(\lambda a)\}][M_0(\lambda a)\, M_1(\lambda a)]$$

$\therefore$ From Eq. (15.63), the multiplying factor

$$= \left(\frac{a\lambda}{2}\right)\left[\frac{M_0(\lambda a)}{M_1(\lambda a)}\right][\sin\{(-\pi/4) + \theta_1(\lambda a) - \theta_0(\lambda a)\}] \text{ and } \lambda a = a\left(\frac{\omega\mu_0\mu_r}{\rho}\right)^{1/2} \tag{15.65}$$

The multiplying factor increases with resistance.

15.6 SKIN EFFECT IN PLATED CONDUCTORS

In radio frequency and higher frequency transmissions, the resonant cavities and waveguides are often silver-plated to utilize the low resistivity of silver, and thus reduce the ohmic losses in these devices. It is of interest to determine the variation of the loss with the thickness of the plated coating. Another type of practical application is that of tinned copper wires. We can study the effects of such coatings on the resistance of such wires.

So once again, we consider a semi-infinite block of metal of resistivity ρ_2, coated with a second metal layer of resistivity ρ_1 and thickness a (Figure 15.13). The flow of an alternating

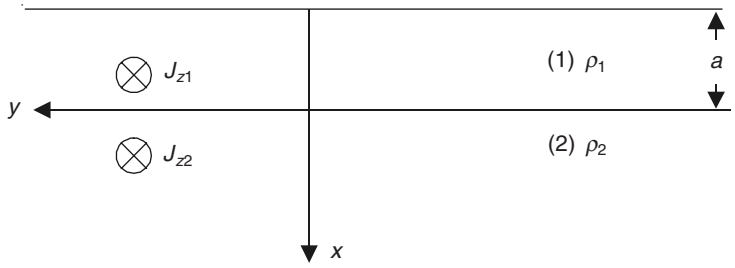


Figure 15.13 Semi-infinite metal block with a surface coating of thickness *a*.

current of frequency $\omega = 2\pi f$ is in the z-direction, and is uniform in the y-direction, i.e. the only variation is in the x-direction. The flow of the current would induce a magnetic field, which would be directed in the y-direction, and also be a function of x. Let the current densities be J_{z1} and J_{z2} in the two media respectively, and as derived in Section 15.2, the equations satisfied by these current density vectors come out to be

$$\frac{d^2J_{z1}}{dx^2} - \left(\frac{j}{d_1^2}\right)J_{z1} = 0 \quad \text{and} \quad \frac{d^2J_{z2}}{dx^2} - \left(\frac{j}{d_2^2}\right)J_{z2} = 0 \tag{15.66}$$

where

$$d_1 = \left(\frac{\rho_1}{\omega\mu_1}\right)^{1/2} \quad \text{and} \quad d_2 = \left(\frac{\rho_2}{\omega\mu_2}\right)^{1/2} \tag{15.67}$$

The general solutions will be

$$J_{z1} = A_1 \exp\left(\frac{x\sqrt{j}}{d_1}\right) + B_1 \exp\left(\frac{-x\sqrt{j}}{d_1}\right) \quad \text{and} \quad J_{z2} = A_2 \exp\left(\frac{x\sqrt{j}}{d_2}\right) + B_2 \exp\left(\frac{-x\sqrt{j}}{d_2}\right)$$

remembering that J_{z1} and J_{z2} are both phasors (i.e. complex vectors with only one component i.e. in the z-direction only).

Since the current density cannot become infinite as $x \to \infty$,

$$J_{z2} = B_2 \exp\left(\frac{-x\sqrt{j}}{d_2}\right) \quad \text{and} \quad J_{z1} = A_1 \exp\left(\frac{x\sqrt{j}}{d_1}\right) + B_1 \exp\left(\frac{-x\sqrt{j}}{d_1}\right) \tag{15.68}$$

The other boundary conditions are at $x = 0$,

$$E_{z1} = E_{z2} \quad \text{and} \quad H_{y1} = H_{y2}$$

∴ From Eqs. (15.48)

$$E_{z1} = \rho_1 J_{z1} = \rho_1 (A_1 + B_1) = \rho_2 B_2 = (\rho_2 J_{z2} = E_{z2}) \tag{15.69}$$

and

$$H_{y1} = \left(\frac{1}{j\omega\mu_1}\right)\left(\frac{dE_{z1}}{dx}\right) = \left(\frac{\rho_1}{j\omega\mu_1}\right)\left(\frac{dJ_{z1}}{dx}\right) = \left(\frac{\rho_1}{j\omega\mu_1}\right)\frac{\sqrt{j}}{d_1}$$

$$= \frac{d_1}{\sqrt{j}}(A_1 - B_1) = \left(\frac{\rho_2}{j\omega\mu_2}\right)\frac{\sqrt{j}}{d_2}(-B_2) = \frac{d_2}{\sqrt{j}}(-B_2) \tag{15.70}$$

We shall now express the current densities in terms of the surface value of J_{z2} on the surface plane $x = 0$, i.e.

$$\text{at } x = 0,\ J_{z2} = B_2 = J_{02} \tag{15.71}$$

Hence from Eqs. (15.69), (15.70), and (15.71), we get

$$J_{z1} = J_{02}\left(\frac{d_2}{\mu_1 d_1^2}\right) \times \left[\left\{\mu_2 d_2 \cosh\left(\frac{x}{d_1\sqrt{2}}\right) - \mu_1 d_1 \sinh\left(\frac{x}{d_1\sqrt{2}}\right)\right\}\cos\left(\frac{x}{d_1\sqrt{2}}\right)\right.$$

$$\left. - j\left\{\mu_1 d_1 \cosh\left(\frac{x}{d_1\sqrt{2}}\right) - \mu_2 d_2 \sinh\left(\frac{x}{d_1\sqrt{2}}\right)\right\}\sin\left(\frac{x}{d_1\sqrt{2}}\right)\right]$$

$$J_{z2} = J_{02}\exp\left(\frac{-x}{d_2\sqrt{2}}\right)\left[\cos\left(\frac{x}{d_2\sqrt{2}}\right) - j\sin\left(\frac{x}{d_2\sqrt{2}}\right)\right] \tag{15.72}$$

It should be noted that there is now a discontinuity in current density at the interface plane between the two metals. So if say a brass bar is coated with silver, then the silver part would carry a larger share of current than if it had been brass in its place.

So now we calculate the total current in each of the two metals for a width, say, W,

$$\therefore \quad I_2 = W\int_0^\infty J_{z2}\,dx = J_{02}W\int_0^\infty \exp\left[\frac{-(1+j)x}{d_2\sqrt{2}}\right]dx = \left(\frac{J_{02}Wd_2}{\sqrt{2}}\right)(1-j)$$

and

$$I_1 = W\int_0^\infty J_{z1}\,dx = \frac{J_{02}Wd_2}{\sqrt{2}\mu_1 d_1}\left[\left(P - \mu_1 d_1\sqrt{2}\right) - j\left(Q - \mu_1 d_1\sqrt{2}\right)\right]$$

where

$$P = \left(\mu_2 d_2\sqrt{2}\right)\times\left[\sinh\left(\frac{a}{d_1\sqrt{2}}\right)\cos\left(\frac{a}{d_1\sqrt{2}}\right) + \cosh\left(\frac{a}{d_1\sqrt{2}}\right)\sin\left(\frac{a}{d_1\sqrt{2}}\right)\right]$$

$$+ \left(\mu_1 d_1\sqrt{2}\right)\times\left[\cosh\left(\frac{a}{d_1\sqrt{2}}\right)\cos\left(\frac{a}{d_1\sqrt{2}}\right) + \sinh\left(\frac{a}{d_1\sqrt{2}}\right)\sin\left(\frac{a}{d_1\sqrt{2}}\right)\right]$$

$$Q = \left(\mu_2 d_2\sqrt{2}\right)\times\left[\sinh\left(\frac{a}{d_1\sqrt{2}}\right)\cos\left(\frac{a}{d_1\sqrt{2}}\right)-\cosh\left(\frac{a}{d_1\sqrt{2}}\right)\sin\left(\frac{a}{d_1\sqrt{2}}\right)\right]$$

$$+\left(\mu_1 d_1\sqrt{2}\right)\times\left[\cosh\left(\frac{a}{d_1\sqrt{2}}\right)\cos\left(\frac{a}{d_1\sqrt{2}}\right)-\sinh\left(\frac{a}{d_1\sqrt{2}}\right)\sin\left(\frac{a}{d_1\sqrt{2}}\right)\right]$$

$\therefore$ The total current, $I = I_1 + I_2$

$$= \left(\frac{J_{02}Wd_2}{\sqrt{2}\mu_1 d_1}\right)(P - jQ) = \left(\frac{J_{02}Wd_2}{\sqrt{2}\mu_1 d_1}\right)\left(P^2+Q^2\right)^{1/2}\angle -\alpha$$

where

$$\alpha = \tan^{-1}\left(\frac{Q}{P}\right) \tag{15.73}$$

We now calculate the power loss in a rectangular section of the plated conductor, of width W in the y-direction and l in the z-direction.

$\therefore$ The loss in the metal 2, for the rectangular section:

$$P_2 = \rho_2 Wl\int_{x=0}^{x\to\infty}\left|J_{z2}\right|^2 dx = \rho_2 WlJ_{02}^2\int_0^{\infty}\exp\left(\frac{x\sqrt{2}}{d_2}\right)dx$$

$$= \frac{\rho_2 Wld_2J_{02}^2}{\sqrt{2}} \tag{15.74}$$

And for the conductor 1,

$$P_1 = \rho_1 Wl\int_{x=0}^{x\to\infty}\left|J_{z1}\right|^2 dx$$

$$= \left(\frac{\rho_1 WlJ_{02}^2d_2^2}{2\sqrt{2}\mu_1^2d_1^3}\right)\times\left[\left(\mu_2 d_2\sqrt{2}\right)^2+\left(\mu_1 d_1\sqrt{2}\right)^2\right]\sinh\left(\frac{a}{d_1\sqrt{2}}\right)\cosh\left(\frac{a}{d_1\sqrt{2}}\right)$$

$$+\left[\left(\mu_2 d_2\sqrt{2}\right)^2-\left(\mu_1 d_1\sqrt{2}\right)^2\right]\sin\left(\frac{a}{d_1\sqrt{2}}\right)\cos\left(\frac{a}{d_1\sqrt{2}}\right)+\mu_1\mu_2 d_1 d_2\sinh^2\left(\frac{a}{d_1\sqrt{2}}\right) \tag{15.75}$$

$\therefore$ The total power $P_T = P_1 + P_2$

$$= \left[\frac{\rho_1 WlJ_{02}^2d_2^2}{\left(2\sqrt{2}\mu_1^2d_1^3\right)}\right]U$$

where

$$U = \left(\frac{1}{2}\right)\left[\left\{\left(\mu_2 d_2\sqrt{2}\right)^2+\left(\mu_1 d_1\sqrt{2}\right)^2\right\}\sinh\left(\frac{a}{d_1\sqrt{2}}\right)\right.$$

$$\left.+\left\{\left(\mu_2 d_2\sqrt{2}\right)^2-\left(\mu_1 d_1\sqrt{2}\right)^2\right\}\sin\left(\frac{a}{d_1\sqrt{2}}\right)+\mu_1\mu_2 d_1 d_2\cosh\left(\frac{a}{d_1\sqrt{2}}\right)\right] \tag{15.76}$$

∴ The equivalent resistance of the composite bar is

$$R_{eq} = \frac{P_T}{|I|^2} = \left(\frac{\rho_1 l}{W d_1 \sqrt{2}}\right)\left(\frac{U}{P^2 + Q^2}\right) \tag{15.77}$$

We now express this equivalent resistance in terms of the resistance of a homogeneous bar of the plating metal, i.e. metal 1,

$$\therefore \qquad R_1 = \frac{\rho_1 l}{W d_1 \sqrt{2}} \qquad \text{and} \qquad \frac{R_{eq}}{R_1} = \frac{U}{P^2 + Q^2} \tag{15.78}$$

And, if we do a similar comparison with the metal 2, then

$$\frac{R_{eq}}{R_2} = \left(\frac{\mu_1 d_1}{\mu_2 d_2}\right)\left(\frac{U}{P^2 + Q^2}\right) \tag{15.79}$$

We now compare the resistance of a silver-plated brass bar with that of a solid bar of the same dimensions, as shown in the Figure 15.14. It is seen that if the silver plate thickness

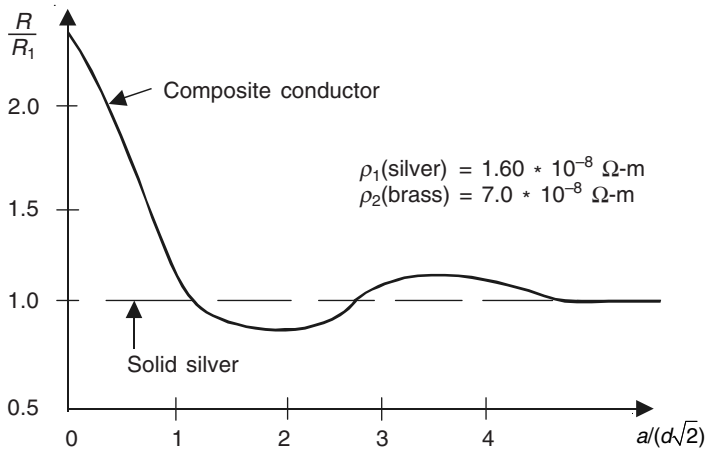


Figure 15.14 Resistance ratio for a silver-plated brass block.

exceeds 1.2 times the skin-depth ($d_1\sqrt{2}$), then the resistance of the composite bar is nearly the same as that of pure silver. The surprising point to be noted is that, there is a dip in this curve from $[a/(d_1\sqrt{2})] = 1.2$ to 2.6, when for this thickness, the total resistance of the composite bar and hence its I^2R current loss as well, are *less than those for a pure silver bar.*

We can similarly analyze other composite metal bars, such as copper–lead, and so on.

15.7 MAGNETIC DIFFUSION AS AN ELECTRICAL TRANSIENT

So far all the problems we have considered are of the type in which the currents have been induced in the conducting media as a consequence of imposing a steady sinusoidally varying

(with time, i.e. time-harmonic) magnetic field on it, or by forcing a similar alternating current on the medium. Also in the present problem under consideration, the material medium is at rest, and we study the diffusion of the magnetic field and the associated induced eddy currents in the conducting medium, when the forcing magnetic field is suddenly switched on as a step-function of time. We devise a model such that we have simplified the geometry with only one component of the magnetic field existing, and one orthogonal component of the electric field and hence the current as well are present in the model, so that the analysis is a simple one (from the mathematical viewpoint). Thus we have an electromagnet of an infinitely permeable, non-conducting magnetic material with an air-gap of length w (Figure 15.15), excited through an N-turn winding by a constant-current source of current I, which can be turned on or off by the switch S. In the gap is a slab of non-magnetic conducting material of dimensions $l \times w \times b$ (l being the length, w the width, and b the thickness) as shown in Figure 15.15, with constant conductivity σ (or resistivity $\rho = 1/\sigma$). The dimensions w and l are large compared with the thickness b, and hence the edge effects (i.e. fringing) can be neglected. The coordinate system for the problem is as defined in Figure 15.15(b). Initially it is assumed that the switch S is

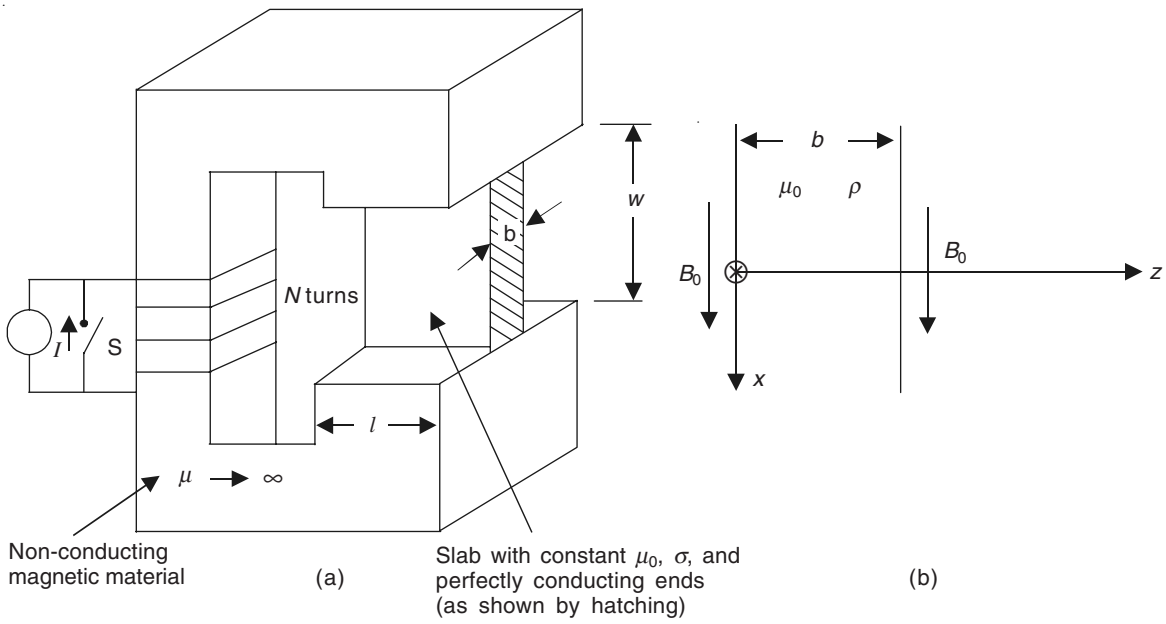


Figure 15.15 (a) System devised to study magnetic diffusion as an electrical transient; (b) coordinate system for the one-dimensional diffusion model.

closed, and hence there is no current in the winding and hence no magnetic flux in the magnet. At the instant $t = 0$, the switch S is opened, and now we shall study the time variations of the flux.

The boundary conditions on opening the switch are:

Outside the block, in the gap of the magnet, for $t > 0$, $z < 0$, $z > b$,

$$\mathbf{B} = \mathbf{i}_x B_0$$

where

$$B_0 = \frac{\mu_0 NI}{w} \tag{15.80}$$

It should be noted that only the x-component of **B** exists in the air-gap of the magnet, and it varies with z only.

Inside the block (of thickness b), the displacement currents can be neglected. The relevant Maxwell's equations are

$$\text{curl } \mathbf{H} = \nabla \times \mathbf{H} = \mathbf{J}$$

and

$$\text{curl } \mathbf{E} = \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$

Since **B** has x-component only and is a function of z only,

$$\text{div } \mathbf{B} = \nabla \cdot \mathbf{B} = 0$$

is automatically satisfied.

The constituent relations are:

$$\mathbf{E} = \rho \mathbf{J} \qquad \text{or} \qquad \mathbf{J} = \sigma \mathbf{E}$$

and

$$\mathbf{B} = \mu_0 \mathbf{H}$$

From the curl **H** equation,

$$\mathbf{i}_y \left(\frac{\partial H_x}{\partial z} \right) = \mathbf{i}_x 0 + \mathbf{i}_y J_y + \mathbf{i}_z 0$$

or

$$\left(\frac{\partial H_x}{\partial z} \right) = \left(\frac{1}{\mu_0} \right) \left(\frac{\partial B_x}{\partial z} \right) = J_y = \sigma E_y$$

$\therefore$ **E** can have y-component only.

$\therefore$ From the curl **E** equation,

$$\mathbf{i}_x \left(-\frac{\partial E_y}{\partial z} \right) = -\mathbf{i}_x \left(\frac{\partial B_x}{\partial t} \right) \qquad \text{or} \qquad \frac{\partial E_y}{\partial z} = \frac{\partial B_x}{\partial t}$$

Combining the above two equations,

$$\left(\frac{1}{\mu_0 \sigma} \right) \left(\frac{\partial^2 B_x}{\partial z^2} \right) = \frac{\partial E_y}{\partial z} = \frac{\partial B_x}{\partial t} \tag{15.81}$$

As soon as the switch S is opened, at that instant,

$$\text{at } t = 0^+, \; B_x = 0 \qquad \text{(inside the conducting block).} \tag{15.82}$$

The distribution of B_x at $t = 0^+$ is shown diagrammatically in the air-gap of the magnet in Figure 15.16.

This condition follows from the equation: $\nabla \times \mathbf{E} = -\dfrac{\partial \mathbf{B}}{\partial t}$

or

$$\oint_C \mathbf{E} \cdot d\mathbf{l} = -\left(\frac{\partial}{\partial t} \right) \iint_S \mathbf{B} \cdot d\mathbf{S} = \frac{1}{\sigma} \oint_C \mathbf{J} \cdot d\mathbf{l} \tag{15.83}$$

which is essentially the Stokes' theorem.

The implication of the ideal B_x distribution as shown by Eq. (15.82) and as shown in Figure 15.16 is that the external flux density in the air-gap (i.e. external to the conducting

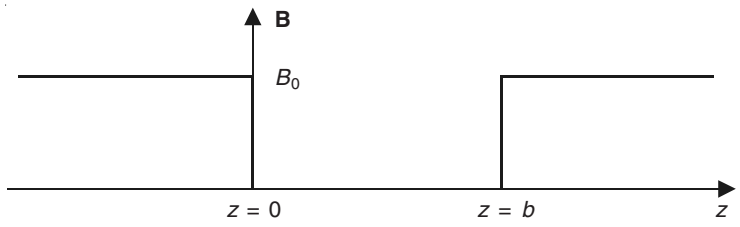


Figure 15.16 Distribution of B_x at $t = 0^+$ in the conducting slab.

block) is established in zero time. This mathematical simplification implies that at this instant, the time-derivative of **B** will be infinite. In the contour C of the Stokes' theorem equation, if C includes $z = 0$ or $z = b$, then it indicates an infinite current density, but actually a finite surface current is necessary to terminate the B_0 field. In a real situation, the flux density B_0 is established rapidly but in a finite time. This idealization will be better understood when we discuss the solution at the completion of this problem.

It should also be understood that after a sufficient long time, all transient currents in the slab will die out, and then B_x will become uniform throughout the block, i.e.

$$\{\text{as } t \to \infty,\ B_x \to B_0 \text{ and then } (\partial B_x/\partial t) = 0 \text{ for the steady-state conditions.}\} \tag{15.84}$$

So to solve Eq. (15.81), we use the method of separation of variables, and assume a solution of the form

$$B_x = B(z)\ B(t) + C_0 \tag{15.85}$$

Substituting in Eq. (15.81),

$$\left[\frac{1}{\mu_0 \sigma B(z)}\right]\left[\frac{d^2 B(z)}{dz^2}\right] = \left[\frac{1}{B(t)}\right]\left[\frac{dB(t)}{dt}\right] = -\alpha$$

(α being an arbitrary constant at this stage).

$$\therefore \qquad B(z) = C_1 \sin\left(\sqrt{\mu_0 \sigma \alpha}\, z\right) + C_2 \cos\left(\sqrt{\mu_0 \sigma \alpha}\, z\right)$$

and

$$B(t) = \exp(-\alpha t)$$

$\therefore$ The solution will have the general form

$$B_x = \left[C_1 \sin\left(\sqrt{\mu_0 \sigma \alpha}\, z\right) + C_2 \cos\left(\sqrt{\mu_0 \sigma \alpha}\, z\right)\right] \exp(-\alpha t) + C_0 \tag{15.86}$$

where C_1, C_2, C_0, and α are constants to be determined from the boundary conditions.

For the necessary initial condition of B_x at $t = 0^+$, from $z = 0$, $C_2 = 0$, and $C_0 = B_0$.

$$\therefore \qquad B_x = C_1 \exp(-\alpha t) \sin\left(\sqrt{\mu_0 \sigma \alpha}\, z\right) + B_0 \tag{15.87}$$

This equation must satisfy the initial conditions of B_x over the range $z = 0$ to $z = b$, as shown in the rectangular distribution represented diagrammatically in Figure 15.16, i.e.

$$\text{for } 0 < z < b, \quad 0 = \sum_{n=1}^{\infty} a_n \sin\left(\frac{n\pi z}{b}\right) + B_0 \tag{15.88}$$

$$\therefore \qquad \sqrt{\mu_0 \sigma \alpha_n} = \frac{n\pi}{b} \quad \text{or} \quad \alpha_n = \frac{n^2\pi^2}{\mu_0 \sigma b^2} \tag{15.89}$$

The above equation has been obtained by expanding the rectangular distribution in a 'Fourier sine series'. So to evaluate a_n, multiplying both sides of the series expansion by sin $(m\pi z/b)$, and integrating over the limits of the periodicity 0 to b, i.e.

$$\int_0^b B_0 \sin\left(\frac{m\pi z}{b}\right) dz = \int_{n=1}^{\infty} \sum a_n \sin\left(\frac{n\pi z}{b}\right) \sin\left(\frac{m\pi z}{b}\right) dz \tag{15.90}$$

All the right-hand terms for which $m \neq n$ vanish, leaving:

$$\int_0^b a_n \sin^2\left(\frac{n\pi z}{b}\right) dz = -B_0 \int_0^b \sin\left(\frac{n\pi z}{b}\right) dz$$

$$\therefore \qquad a_n = -\left(\frac{4}{n\pi}\right) B_0 \qquad \text{for } n = 1, 3, 5, \ldots$$

$$= 0 \qquad \text{for } n = 2, 4, 6, \ldots \tag{15.91}$$

$$\therefore \qquad B_x = B_0 \left[1 - \sum_{n=1,3,5,\ldots}^{\infty} \left(\frac{4}{n\pi}\right) \{\exp(-\alpha_n t)\} \sin\left(\frac{n\pi z}{b}\right)\right] \tag{15.92}$$

Note that each space harmonic damps at different rate, the higher harmonics damping faster.

We now define the fundamental time constant τ as

$$\tau = \frac{1}{\alpha_1} = \frac{\mu_0 \sigma b^2}{\pi^2}, \qquad n = 1 \tag{15.93}$$

$$\therefore \qquad B_x = B_0 \left[1 - \sum_{n=1,3,5,\ldots}^{\infty} \left(\frac{4}{n\pi}\right) \left\{\exp\left(\frac{-n^2 t}{\tau}\right)\right\} \sin\left(\frac{n\pi z}{b}\right)\right] \tag{15.94}$$

The fundamental time constant, which is the longest time-constant of the series, is called the 'diffusion time-constant' of the system.

We now plot the flux density distribution as a function of z in the thickness of the block b at different instants of time (Figure 15.17). It will be noted that, at $t = 0$, the flux is completely eliminated from the slab. With the progress of time, the flux diffuses into the slab. Because the higher harmonics damp out faster,

at $t = 0.1\tau$, only 3 terms are needed for B_x calculation;
at $t = 0.3\tau$, only 2 terms are required; and
for $t > \tau$, only the fundamental term is significant.

This shows why the fundamental diffusion time τ is the controlling time constant in the diffusion process. Also, the field is nearly completely diffused in the material in time, $t > 3\tau$.

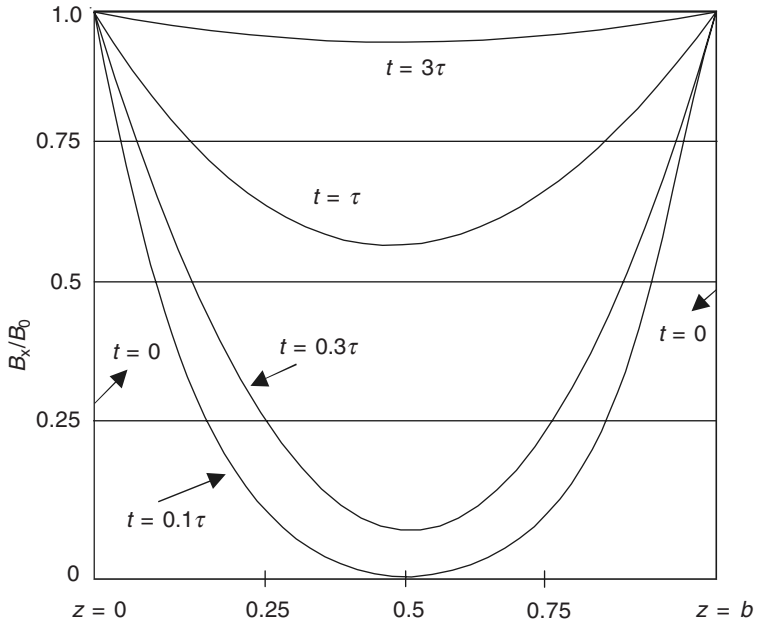


Figure 15.17 B_x as a function of z, with the time t as a parameter.

Next, we consider the behaviour of the current density inside the slab during the transient diffusion process. The current density is given by

$$J_y = \left(\frac{1}{\mu_0}\right)\left(\frac{\partial B_x}{\partial z}\right) = -\left(\frac{B_0}{\mu_0 b}\right)\sum_{n=1,3,5,\cdots}^{\infty}\left[4\left\{\exp\left(\frac{-n^2 t}{\tau}\right)\right\}\cos\left(\frac{n\pi z}{b}\right)\right] \tag{15.95}$$

The space distribution of current density J_y is shown in the Figure 15.18. Note from this equation, that at $t = 0$, this series does not converge at $z = 0$ and $z = b$. This shows that at $t = 0$, the magnetic field is completely excluded from the slab, and to terminate the external magnetic field, a surface current density is required—which implies an infinite volume current density—thus at $t = 0$, there are impulses of J_y at $z = 0$ and $z = b$. As the time progresses, the currents diffuse in the slab and decay. At the surface of the slab (i.e. $z = 0$ and $z = b$) the current density decays continually, but at an interior point, the current density increases from zero to some maximum value and then decays back to zero.

Conservation of current requires that the current flows down one side of the block and return on the opposite side. Hence in Figure 15.18, the current distribution graphs show an odd symmetry (or skew symmetry) about $z = b/2$. Note also that a one-dimensional solution implies perfectly conducting end-plates on the slab to provide a return path for the currents.

Notes:

1. It is a point to be noted that for such electromagnetic phenomena, occurring outside the slab, which have characteristic times much shorter than the diffusion time τ, the properties of the slab can be approximated by assuming its $\sigma \to \infty$. This is a valid assumption, used in calculating the sub-transient component of the end-winding reactances in large turbogenerators, i.e. the time-duration is so short that only the air-

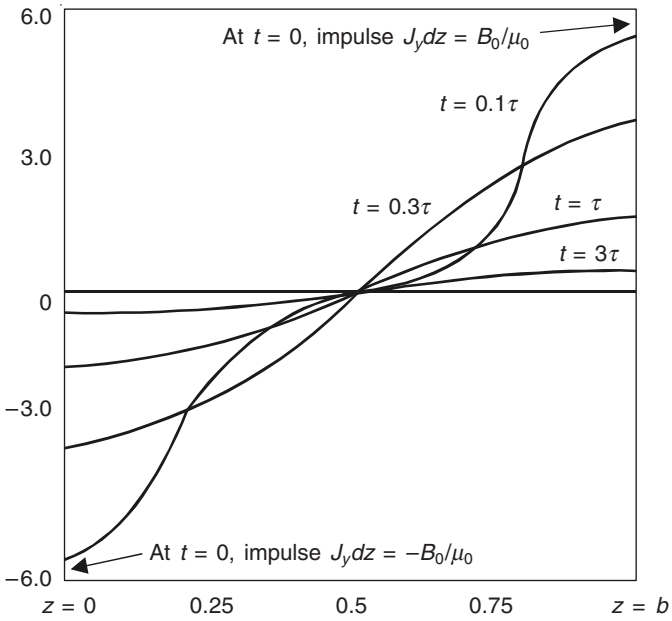


Figure 15.18 Current distribution in one-dimension diffusion.

region flux has been established and the flux has been excluded from the interior of the block (which in the case of the turbogenerator end-region is the clamping plate on the core-end-surface).

2. On the other hand, for those phenomena which have characteristic times much longer than the diffusion time τ, the perturbations due to the diffusion effects can be neglected for field calculations. These two limiting cases are analogous to the constant current and the constant flux conditions of the lumped parameter systems.

Some examples of diffusion time constants ($\tau = \mu_0\mu_r\sigma b^2/\pi^2$) are:

1. A slab of copper 1 cm thick, $\sigma = 5.9 \times 10^7$ mhos/m; $\tau = 1$ ms
2. Seeded combustion gas, 1 metre thick, $\sigma = 40$ mhos/m; $\tau = 10$ μs
3. Silicon-iron (4%), 1 cm thick, $\sigma = 1.7 \times 10^6$ mhos/m, $\mu_r = 5000$, $\tau = 100$ ms

We see that even though the conductivity of Si-Fe is much lower than that of copper, its diffusion time is much higher because of high relative permeability. This high diffusion time constant of Si–Fe makes the laminations of the iron cores necessary in all alternating current equipment, such as generators, motors, and transformers.

15.8 MAGNETIC DIFFUSION AS A RESULT OF RELATIVE MOTION OF THE CONDUCTING MEDIUM

So far we have considered the phenomenon of the magnetic diffusion caused by changes in the magnetic field with time, i.e. either due to a transient change in the field or due to a change in the field in a regular periodic manner with time (time-harmonic fields). In all these cases, it has been assumed that there has been no relative motion between the conducting body and the source field. However, it is also possible to produce the diffusion effect by a relative

displacement between the source field and the conducting body (taking place in a finite time). Such continuous displacements can be produced either by a moving conducting body or by a travelling magnetic field pattern or by a combination of both. Before we study specific examples of such a phenomenon, we shall extend the operating equation [i.e. Eq. (12.39)] to include the effects of the relative motion. This can be done by either of the two ways.

Maxwell's equation of electromagnetic induction (or Faraday's law of induction) which has been used to derive Eq. (12.39), does not include the effects of the moving media, even though we have generalized this equation [i.e. Eq. (12.10) in integral form, or Eq. (12.17) in differential form] to include such effects in Chapter 10, Sections 10.3 to 10.4. So now we use this generalized equation along with other relevant equations to derive the generalized diffusion equation. For this purpose, our relevant Maxwell's equations are

$$\nabla \cdot \mathbf{B} = 0 \tag{12.16}$$

$$\nabla \times \mathbf{H} = \mathbf{J} \tag{12.18′}$$

$$\nabla \times \mathbf{E} = -\left(\frac{\partial \mathbf{B}}{\partial t}\right) + \nabla \times (\mathbf{v} \times \mathbf{B}) \tag{10.29′}$$

Here Eq. (12.18′) is the Ampere's law Eq. (12.18) in which the displacement current term ($\partial \mathbf{D}/\partial t$) has been neglected; and Eq. (10.29′) is the differential form of the generalized electromagnetic induction Eq. (10.29), which was written in the integral form. (And $\mathbf{v}$ is the velocity of the moving medium.)

The relevant constituent relations for this phenomenon are

$$\mathbf{B} = \mu \mathbf{H} \tag{12.13}$$

$$\mathbf{E} = \rho \mathbf{J} \quad \text{or} \quad \mathbf{J} = \sigma \mathbf{E} \tag{12.14}$$

Rewriting Eq. (12.17′) with Eq. (12.14), we get

$$\nabla \times \mathbf{H} = \sigma \mathbf{E}$$

and doing a curl operation on this equation, we get

$$\left(\frac{1}{\sigma}\right)\nabla \times \nabla \times \mathbf{H} = \nabla \times \mathbf{E} = -\left(\frac{\partial \mathbf{B}}{\partial t}\right) + \nabla \times (\mathbf{v} \times \mathbf{B})$$

from Eq. (10.29′).

Converting this equation to a single variable (i.e. $\mathbf{B}$) equation by using Eq. (12.13), we get

$$\left(\frac{1}{\mu\sigma}\right)\nabla \times \nabla \times \mathbf{B} = -\left(\frac{\partial \mathbf{B}}{\partial t}\right) + \nabla \times (\mathbf{v} \times \mathbf{B}) \tag{15.96}$$

The left-hand side of this equation can be expressed as

$$-\left(\frac{1}{\mu\sigma}\right)\nabla \times \nabla \times \mathbf{B} = -\left(\frac{1}{\mu\sigma}\right)\left[\nabla(\nabla \cdot \mathbf{B}) - \nabla^2 \mathbf{B}\right] \text{—vector identity}$$

and using Eq. (12.16) which makes the first term of this equation to vanish, we get

$$-\left(\frac{1}{\mu\sigma}\right)\nabla^2 \mathbf{B} + \left(\frac{\partial \mathbf{B}}{\partial t}\right) = \nabla \times (\mathbf{v} \times \mathbf{B}) \tag{15.97}$$

This equation describes the distribution of the magnetic field in the conducting medium. It includes both the effects of the time-varying magnetic field and the material motion, and can be used for solving a variety of practical engineering problems. This equation is also of great importance in determining the magnetic field origins in the liquid core of the earth, and in this context it is sometimes referred to as the 'Bullard's equation'. This equation can also be derived in an alternative manner, using the restricted equation for the electromagnetic induction, i.e. Eq. (12.18) and then replacing the constituent relation (12.14) by a more general form in which the effects of relative motion are included. We shall see later in Chapter 20, when we deal with moving coordinate systems (Electromagnetism and Special Relativity), that the current density vector **J** can be expressed as

$$\mathbf{J} = \sigma(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \tag{15.98}$$

This equation along with the other constituent relation

$$\mathbf{B} = \mu\mathbf{H} \tag{12.13}$$

and the relevant Maxwell's equations

$$\nabla \cdot \mathbf{B} = 0 \tag{12.16}$$

$$\nabla \times \mathbf{H} = \mathbf{J} \tag{12.18′}$$

$$\nabla \times \mathbf{E} = -\left(\frac{\partial \mathbf{B}}{\partial t}\right) \tag{12.17}$$

would give us the same result (as we have seen in the earlier part of this section).

We now start with Eq. (12.17), and obtain

$$\nabla \times \mathbf{E} = \nabla \times \left[\left(\frac{1}{\sigma}\right)\mathbf{J} - \mathbf{v} \times \mathbf{B}\right] = -\left(\frac{\partial \mathbf{B}}{\partial t}\right)$$

by using Eq. (15.98).

Or

$$\left(\frac{1}{\sigma}\right)\nabla \times \mathbf{J} + \frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B})$$

or

$$\left(\frac{1}{\mu\sigma}\right)\nabla \times \nabla \times \mathbf{B} + \frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B})$$

by using Eq. (12.18′).

And, as before, the operator identity gives us

$$\nabla \times \nabla \times \mathbf{B} = [\nabla(\nabla \cdot \mathbf{B}) - \nabla^2\mathbf{B}] = -\nabla^2\mathbf{B} \text{ [from Eq. (12.16)].}$$

∴ The above equation reduces to

$$-\left(\frac{1}{\mu\sigma}\right)\nabla^2\mathbf{B} + \frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) \tag{15.97′}$$

which is same as Eq. (15.97), i.e. the 'general equation for magnetic diffusion' which includes both the effects of time-varying magnetic fields as well as material motion.

There are again two limiting cases:

1. When the material motion is absent, the equation simplifies to

$$\left(\frac{1}{\mu\sigma}\right)\nabla^2\mathbf{B} = \frac{\partial\mathbf{B}}{\partial t} \tag{15.99}$$

which is really the vector diffusion equation of (12.39) in which the time-variation operator $(\partial/\partial t)$ has been kept in the general form, and not restricted to the time-harmonic variation. In fact, it is this equation, which we have been solving in the problems discussed in Sections 15.2 to 15.6.

2. The second limiting case, which is also of considerable importance, is also obtained when there is only motion under steady-state conditions, i.e. the magnetic field vector is not time-varying. We shall look at some situations consistent with this condition. The equation for this limiting condition is

$$-\left(\frac{1}{\mu\sigma}\right)\nabla^2\mathbf{B} = \nabla \times (\mathbf{v} \times \mathbf{B}) \tag{15.99′}$$

However some of the most important engineering examples are those of the time-varying excitation in presence of material motion. The induction motors are examples of interaction of this type. Now we shall have a look at some of these problems in a simplified manner.

15.9 STEADY-STATE MAGNETIC FIELD IN A FIXED FRAME

This type of steady-state (non-time-varying) diffusion problem is relevant to magneto-hydrodynamic (MHD) generators. The system consists of a continuous strip of material of constants, σ, ε_0, μ_0, sliding with constant velocity $\mathbf{v}$ between a parallel pair of highly conducting electrodes. The dimensions and the coordinate system are as shown in Figure 15.19.

The sliding strip is assumed to make perfect contacts with the electrodes. The system is excited by current sources at the end $z = l$, so that the total current to the electrodes is I amps.

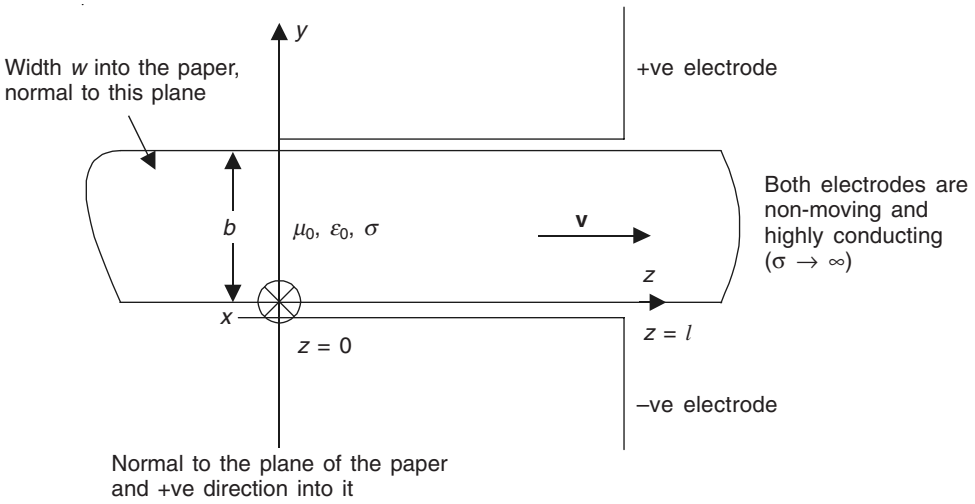


Figure 15.19 Geometry for diffusion in presence of steady motion only.

We intend to determine the **B** distribution in the system, and the effects of the system parameters on this distribution.

Once again, l/b and w/b are assumed to be large enough so that the edge and the end effects can be neglected. The flux density **B** can be taken to be one-dimensional, i.e. only B_x exists, since the current flow is in the y-direction:

$$\mathbf{B} = \mathbf{i}_x B_x,\ \mathbf{J} = \mathbf{i}_y J_y, \qquad \text{and hence } \mathbf{E} = \mathbf{i}_y E_y \tag{15.100}$$

and hence the field quantities vary as a function of z only, and there is no variation with respect to x or y.

Also, the constant velocity of the strip is

$$\mathbf{v} = \mathbf{i}_z \mathrm{v}_z \tag{15.101}$$

From the $\nabla \times \mathbf{E}$ Eq. (10.29′), which now simplifies to $\nabla \times \mathbf{E} = \nabla \times (\mathbf{v} \times \mathbf{B})$, we get

$$\mathbf{i}_x \left(-\frac{\partial E_y}{\partial z} \right) = \mathbf{i}_x \left[-\frac{\partial}{\partial z} \left(\mathrm{v}_z B_x \right) \right]$$

or

$$\left(\frac{1}{\sigma} \right) \left(\frac{\partial J_y}{\partial z} \right) = \mathrm{v}_z \left(\frac{\partial B_x}{\partial z} \right) \tag{15.102}$$

And from the $\nabla \times \mathbf{H}$ Eq. (12.18′):

$$\mathbf{i}_y \left(+\frac{\partial H_x}{\partial z} \right) = \mathbf{i}_y J_y \qquad \text{or} \qquad \left(\frac{1}{\mu_0} \right) \left(\frac{\partial B_x}{\partial z} \right) = J_y \tag{15.103}$$

From Eq. (15.103),

$$\left(\frac{1}{\mu_0} \right) \left(\frac{\partial^2 B_x}{\partial z^2} \right) = \frac{\partial J_y}{\partial z} = \mathrm{v}_z \sigma \left(\frac{\partial B_x}{\partial z} \right)$$

or

$$\left(\frac{d^2 B_x}{dz^2} \right) - \mu_0 \sigma \mathrm{v}_z \left(\frac{dB_x}{dz} \right) = 0 \tag{15.104}$$

The total derivative has replaced the partial derivative, since B_x is a function of z only. Thus the above Eq. (15.104) is the equivalent of Eq. (15.99) for this problem.

The relevant boundary conditions for the problem are:

$$\text{At} \qquad z = 0,\ B_x = 0; \qquad \text{and} \qquad \text{at } z = l,\ B_x = \frac{\mu_0 I}{w} \tag{15.105}$$

When the velocity v_z is a finite constant, the solution of B_x is of the form

$$B_x = C \exp(\alpha z) \tag{15.106}$$

Substituting in Eq. (15.104), we get

$$\alpha^2 - \mu_0 \sigma \mathrm{v}_z \alpha = 0$$

The solutions of this equation are

$$\alpha = \mu_0 \sigma \mathrm{v}_z \text{ and } \alpha = 0$$

$\therefore$ The general solution for B_x is

$$B_x = C_0 + C_1 \exp(\mu_0 \sigma v_z z)$$

Using the boundary conditions of (15.105),

$0 = C_0 + C_1$, from the $z = 0$ condition $\rightarrow C_0 = -C_1$, and $\dfrac{\mu_0 I}{w} = C_0 + C_1 \exp(\mu_0 \sigma v_z l)$

$$\therefore \quad C_1 = \left(\frac{\mu_0 I}{w}\right)\left[\exp(\mu_0 \sigma v_z l) - 1\right]^{-1}$$

We write

$$\mu_0 \sigma v_z l = R_m \tag{15.107}$$

which is dimensionless and defined as the 'magnetic Reynold's number'.

$$\therefore \quad B_x = \left(\frac{\mu_0 I}{w}\right)\frac{\exp\left(\dfrac{R_m z}{l}\right) - 1}{\exp(R_m) - 1} \tag{15.108}$$

and the current density distribution is

$$J_y = \left(\frac{1}{\mu_0}\right)\left(\frac{dB_x}{dz}\right) = \frac{I}{wl}\frac{R_m \exp\left(\dfrac{R_m z}{l}\right)}{\exp(R_m) - 1} \tag{15.109}$$

When $R_m \rightarrow 0$, i.e. the velocity v_z is made zero, these two distributions become

$$B_x = \left(\frac{\mu_0 I}{w}\right)\left(\frac{z}{l}\right) \quad \text{and} \quad J_y = \frac{I}{wl}$$

which can also be obtained by directly putting $v_z = 0$ in the equations (15.104).

We now plot B_x and J_y as functions of position z in the slab for different values of the magnetic Reynold's number R_m (Figure 15.20). It is seen that at $v_z = 0$, the distributions of **B** and **J** in the slab are linear, and as R_m keeps on increasing these distributions become more and more nonlinear due to the higher exponential variations with R_m and velocity. It should also be noted that the direction of motion and the direction of diffusion (of **B** and **J**) are opposite. The magnitude of R_m is an indicator of the relative effectiveness of the two processes. A large R_m produces a slow diffusion. The magnetic Reynold's number can be taken as proportional to the ratio of the 'diffusion time constant' [defined earlier in the Section 15.7, Eq. (15.93)] to the time taken by a sample of the material to traverse the length of the electrodes.

Remembering

$$\tau \text{ or } \tau_d = \frac{\mu_0 \sigma l^2}{\pi^2} \tag{15.93}$$

and the time required by a sample of material to traverse the length l with a velocity v_z is

$$\tau_t = \frac{l}{v_z}$$

$$\therefore \quad \frac{\tau_d}{\tau_t} = \frac{\mu_0 \sigma v_z l}{\pi^2} = \frac{R_m}{\pi^2} \tag{15.110}$$

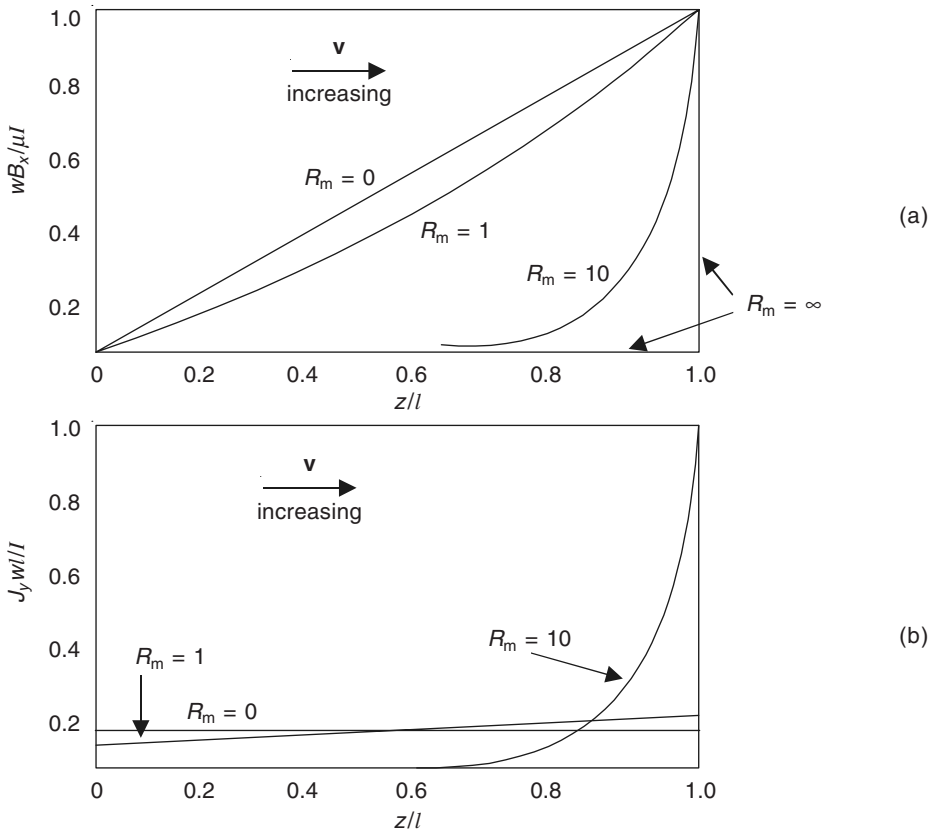


Figure 15.20 *B* and *J* distributions in the system of Figure 15.19 as functions of *z* and R_m: (a) B_x distribution and (b) J_y distribution.

It should be noted that there are other ways of defining R_m as well, which we shall use in other problems.

15.10 SINUSOIDALLY TIME-VARYING MAGNETIC FIELD IN PRESENCE OF MOTION

Now we consider a problem in which we introduce both the characteristic dynamical times into a single system, i.e. a sheet of conducting metal of conductivity σ is moving to the right with a constant velocity **v** in the *z*-direction (Figure 15.21). The sheet is assumed to move in the air-gap of an infinitely permeable magnetic circuit, so that the problem can be considered as a one-dimensional problem. **B** is uniform in the *x*- and *y*-directions, with variations in the *z*-direction, and has only one component, i.e.

$$\mathbf{B} = \mathbf{i}_y \, \mathrm{Re}\left[\hat{B}_0 \exp(j\omega t) \right] \qquad \text{at } z = 0 \tag{15.111}$$

We consider an ideal distribution of this type, without going into the device of its generation, just as we had assumed ideal field distributions in the problems described in Sections 15.2 and 15.3. Since there are no variations in the *x*- and the *y*-directions, the operators

$$\frac{\partial}{\partial x} = \frac{\partial}{\partial y} = 0$$

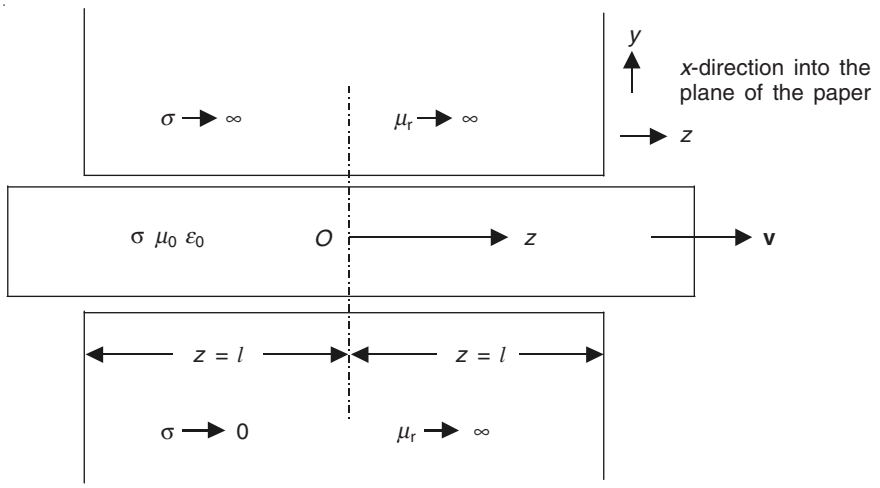


Figure 15.21 A thin slab of conductor moving in the +*z*-direction in the air-gap of a magnetic circuit.

and the one-dimensional magnetic field in the moving sheet then satisfies Eq. (15.97), which in this case simplifies from

$$-\left(\frac{1}{\mu\sigma}\right)\nabla^2\mathbf{B}+\left(\frac{\partial\mathbf{B}}{\partial t}\right)=\nabla\times(\mathbf{v}\times\mathbf{B}) \tag{15.97}$$

to

$$\left(\frac{1}{\mu\sigma}\right)\left(\frac{\partial^2 B_y}{\partial z^2}\right)=\left(\frac{\partial B_y}{\partial t}\right)+\mathrm{v}\left(\frac{\partial B_y}{\partial z}\right) \tag{15.112}$$

and the currents in the slab would flow in the *x*-direction, which would be given by the curl **H** = **J** equation, which in this case simplifies to

$$J_x=-\left(\frac{1}{\mu}\right)\left(\frac{\partial B_y}{\partial z}\right) \tag{15.113}$$

It should be noted that the direction of **B** (the *y*-direction) is at right angles to the direction of motion (the *z*-direction), which is also the direction in which the variations take place.

Since the excitation is sinusoidal with time, we can assume the solution of the equation to be of the form

$$B_y=\mathrm{Re}\left[\hat{B}_y(z)\exp(j\omega t)\right] \tag{15.114}$$

where ω the angular frequency is the same as that of the source excitation. Substituting this expression in Eq. (15.112), we get

$$\frac{\partial^2\hat{B}_y}{\partial z^2}-\mu\sigma\mathrm{v}\left(\frac{\partial\hat{B}y}{\partial z}\right)-j\omega\mu\sigma\hat{B}_y=0 \tag{15.115}$$

The solution of this equation will be of the form

$$\hat{B}_y(z)=\exp(-jkz)$$

Then, this gives (also called the 'dispersion equation'):

$$-k^2 + jk\mu\sigma \mathrm{v} - j\omega\mu\sigma = 0 \tag{15.116}$$

$$\therefore \qquad k = \frac{j\mu\sigma \mathrm{v} \pm \left[-(\mu\sigma \mathrm{v})^2 - 4j\omega\mu\sigma\right]^{1/2}}{2}$$

or normalizing this by using the length l,

$$kl = \frac{j\mu\sigma \mathrm{v}l \pm \left[-(\mu\sigma \mathrm{v}l)^2 - 4j\omega\mu\sigma l^2\right]^{1/2}}{2}$$

We write the magnetic Reynold's number as $R_\mathrm{m} = \mu\sigma \mathrm{v}l$, and the skin-depth as $d\sqrt{2}$, where $d^2 = 1/\omega\mu\sigma$.

$$\therefore \qquad kl = \frac{jR_\mathrm{m}}{2} \pm j\left[\left(\frac{R_\mathrm{m}}{2}\right)^2 + 2j\left(\frac{l}{d\sqrt{2}}\right)\right]^{1/2} \tag{15.117}$$

so that the solution (15.114) takes the form

$$B_y(z, t) = \mathrm{Re}\left[\hat{B}_0 \exp\{j(\omega t - kz)\}\right] \tag{15.118}$$

which is that of a travelling waveform in which the wave number k is complex (in the complex variable sense). Since k is complex, it means that there is a real part of the exponent exp $(-jkz)$ = exp $(-k_i z)$ exp $(-jk_r z)$, where k can be expressed as $k = k_r + jk_i \rightarrow$ exp $(-k_i z)$ shows the attenuating nature of the wave as it progresses in the z-direction—similar to what we have seen in the problem dealt with in Section 15.2 (when R_m is made zero, or v = 0, such that $R_\mathrm{m} = 0$). The attenuation of the dispersion wave (as it is called), then depends wholly on the skin-depth $(= d\sqrt{2})$. In the present problem, the attenuation of the wave is a function of both the skin-depth as well as the magnetic Reynold's number R_m. We shall now have a look at the complex nature of k (the wave number) in the present problem. Equation (15.117) shows that k has two values, which we denote by k^+ and k^-, then both these can be expressed as

$$k^+ = k_r^+ + jk_i^+ \qquad \text{and} \qquad k^- = k_r^- + jk_i^-$$

On separating the real and the imaginary parts of k^+ and k^-, we find that:

$$\text{for } z > 0,\ B_y = \mathrm{Re}\left[\hat{B}_0 \exp\left(-\left|k_i^+\right| z\right) \cdot \exp\left(j\omega t - \left|k_r^+\right| z\right)\right]$$

and

$$\text{for } z < 0,\ B_y = \mathrm{Re}\left[\hat{B}_0 \exp\left(-\left|k_i^-\right| z\right) \cdot \exp\left(j\omega t + \left|k_r^-\right| z\right)\right] \tag{15.119}$$

There are two diffusion waves propagating respectively in the $+z$ and $-z$ directions starting from $z = 0$, with phase velocities ω/k_r^+ and ω/k_r^- respectively, both of which are modified by the material motion. It should be noted that on separating the real and the imaginary parts of k^+ and k^-, we get

$$k_r^- = -k_r^+ \qquad \text{and} \qquad k_i^- > 0 > k_i^+ \tag{15.120}$$

Thus the spatial rate of attenuation of the wave travelling (or dispersing) to the right is lessened due to the motion of the slab also to the right, whereas the wave travelling to the left gets attenuated faster as a consequence of the slab motion in the opposite direction.

A practical application of the magnetic diffusion phenomenon discussed in this section, is that the device shown in Figure 15.21 can be used to measure the velocity of the material. A mechanism based on this phenomenon, for measuring the velocity of the material, has the advantage that it (the mechanism) does not require any mechanical or electrical contact with the moving medium.

15.11 TRAVELLING WAVE DIFFUSION IN MOVING MEDIUM

This problem and the next two are of great practical importance in solving both the operational as well as the design problems of electrical machines (i.e. large induction motors and turbo-generators). However we shall solve here the basic idealized problems (which have closed-form solutions), which can then be extrapolated to match the detailed boundary conditions of the actual practical problems.

The first problem we consider is that of a travelling wave of a surface current sheet (which in an alternating current machine is produced by a uniformly distributed three-phase winding), imposed on a semi-infinite conducting medium (which represents in developed Cartesian geometry, a solid rotor of large ac machines). It (the medium) is subjected to a travelling mmf wave, or preferably a travelling current sheet wave [A exp $\{j(\omega_S t - ky)\}$], the direction of the current flow being in the z-direction, and the sheet lying in the y-z plane on the $x = 0$ surface of the conducting block (Figure 15.22).

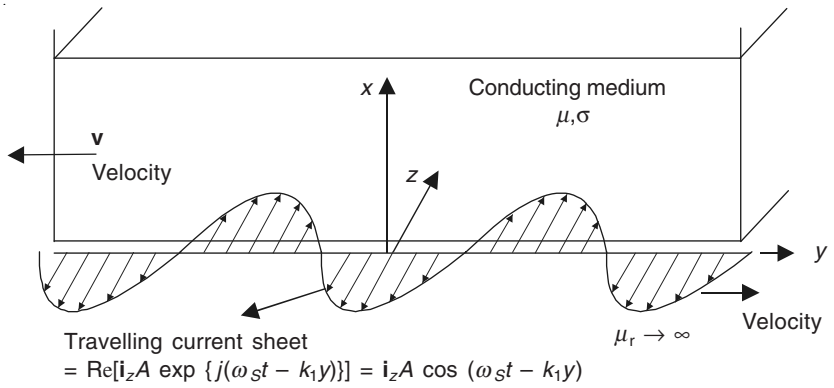


Figure 15.22 Moving conducting medium, subjected to a travelling current sheet.

The practical problem from which this simplified system has been derived, is from the solid rotor surface of a large turbogenerator in which the three-phase currents of the distributed stator winding produce the travelling (or the rotating) current sheet which travels at the synchronous speed of $\omega_S = 2\pi f$ or $\omega_S = 2\pi n_S$ revs/sec for a two-pole machine, or more generally $\omega_S = 2\pi n_S p$ for a $(2p)$-pole machine (n_S, being the number of rotations per second). Hence the current sheet can be expressed as

$$\text{the travelling current sheet} = \mathbf{i}_z \text{Re}[A \exp\{j(\omega_S t - k_1 y)\}], \text{ A real,} \qquad (15.121a)$$

where the wave number $k_1 = (2\pi/\lambda)$, λ being $= \pi D_R$, where D_R is the diameter of the rotor of the machine. The conducting block is moving at the speed $v = n_S \pi D_R$, in the direction $(-y)$, opposite to that of the travelling current sheet; i.e.

$$\mathbf{v} = -\mathbf{i}_y v = -\mathbf{i}_y (n_S \pi D_R) \qquad (15.121b)$$

This model now simulates (in a highly simplified and idealized manner in Cartesian geometry), the problem of a turbogenerator rotor travelling at synchronous speed in one direction, and the negative sequence current sheet travelling at the same synchronous speed in the opposite direction. (In our present analysis, we have taken the current sheet to travel in the forward (or positive) direction, and the rotor block to travel in the opposite (negative or backward) direction; though in an actual turbogenerator, the two travel in the opposite senses. However the final results in loss calculations are not affected by either convention.)

In this case, it will be seen that the components of the magnetic field vector **B** as well as the induced current density vector **J** in the rotor material, would satisfy the same operating equation, which we have derived earlier as

$$\nabla^2 \mathbf{B} = \mu\sigma\left[\left(\frac{\partial \mathbf{B}}{\partial t}\right) - \nabla \times (\mathbf{v} \times \mathbf{B})\right] \tag{15.97}$$

It will also be seen that this problem is no longer a simple one-dimensional problem of the types, which we have been discussing so far in this chapter. The field vectors vary with more than one space variable, as well as the magnetic field vector (i.e. one of the dependent variable vectors) has more than one component. Hence, this being the first problem of relatively more complex nature, we shall, instead of substituting and using directly Eq. (15.97), derive the scalar components of this equation directly from the first principles, i.e. the generalized relevant equations of Maxwell, which for this problem are

$$\nabla \times \mathbf{H} = \mathbf{J} \tag{15.122a}$$

$$\nabla \times \mathbf{E} = -\left(\frac{\partial \mathbf{B}}{\partial t}\right) + \nabla \times (\mathbf{v} \times \mathbf{B}) \tag{15.122b}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{15.122c}$$

and the relevant constituent relations are

$$\sigma \mathbf{E} = \mathbf{J} \tag{15.123a}$$

$$\mathbf{B} = \mu \mathbf{H} \tag{15.123b}$$

and the velocity of the moving block is

$$\mathbf{v} = \mathbf{i}_y(-\mathrm{v}) \tag{15.124}$$

In the medium under consideration, only J_z exists, and there is no variation in the z-direction, i.e. $\partial/\partial z = 0$.

$\therefore$ From Eq. (15.123a),

$$\mathbf{J} = \mathbf{i}_z J_z = \sigma \mathbf{E} = \mathbf{i}_z \sigma E_z \qquad \text{or} \qquad J_z = \sigma E_z \tag{15.125}$$

and from Eq. (15.122a),

$$\nabla \times \mathbf{H} = \mathbf{i}_x 0 + \mathbf{i}_y 0 + \mathbf{i}_z \left(\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y}\right) = \mathbf{i}_z J_z$$

$$\therefore \qquad \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = J_z$$

or

$$\frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y} = \mu\sigma E_z \tag{15.126}$$

and hence from Eq. (15.122b),

$$\mathbf{B} = \mathbf{i}_x B_x + \mathbf{i}_y B_y = \mu(\mathbf{i}_x H_x + \mathbf{i}_y H_y), \qquad H_z = 0$$

$$\therefore \qquad (\mathbf{v} \times \mathbf{B}) = -\,\mathbf{i}_z[(-\,\mathrm{v})(B_x)] = \mathbf{i}_z \mathrm{v} B_x \tag{15.127}$$

$\therefore$ From Eq. (15.122b),

$$\nabla \times \mathbf{E} = \mathbf{i}_x\left(\frac{\partial E_z}{\partial y}\right) + \mathbf{i}_y\left(-\frac{\partial E_z}{\partial x}\right)$$

$$= -\left[\mathbf{i}_x\left(\frac{\partial B_x}{\partial t}\right) + \mathbf{i}_y\left(\frac{\partial B_y}{\partial t}\right)\right] + \mathbf{i}_x\left[\frac{\partial}{\partial y}(\mathrm{v}B_x)\right] + \mathbf{i}_y\left[-\frac{\partial}{\partial x}(\mathrm{v}B_x)\right]$$

This equation gives the following two scalar equations:

$$\frac{\partial E_z}{\partial y} = -\left(\frac{\partial B_x}{\partial t}\right) + \mathrm{v}\left(\frac{\partial B_x}{\partial y}\right)$$

and

$$-\left(\frac{\partial E_z}{\partial x}\right) = -\left(\frac{\partial B_y}{\partial t}\right) - \mathrm{v}\left(\frac{\partial B_x}{\partial x}\right) \tag{15.128}$$

On substituting from Eq. (15.126), these equations become:

$$\left(\frac{1}{\mu\sigma}\right)\left(\frac{\partial^2 B_y}{\partial x \partial y} - \frac{\partial^2 B_x}{\partial y^2}\right) = -\left(\frac{\partial B_x}{\partial t}\right) + \mathrm{v}\left(\frac{\partial B_x}{\partial y}\right)$$

and

$$\left(\frac{1}{\mu\sigma}\right)\left(-\frac{\partial^2 B_y}{\partial x^2} + \frac{\partial^2 B_x}{\partial x \partial y}\right) = -\left(\frac{\partial B_y}{\partial t}\right) - \mathrm{v}\left(\frac{\partial B_x}{\partial x}\right)$$

Equation (15.122c) becomes

$$\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} = 0 \tag{15.129}$$

which combined with the previous equations gives

$$\frac{\partial^2 B_x}{\partial x^2} + \frac{\partial^2 B_x}{\partial y^2} = \mu\sigma\left(\frac{\partial B_x}{\partial t}\right) - \mu\sigma\mathrm{v}\left(\frac{\partial B_x}{\partial y}\right)$$

and

$$\frac{\partial^2 B_y}{\partial x^2} + \frac{\partial^2 B_y}{\partial y^2} = \mu\sigma\left(\frac{\partial B_y}{\partial t}\right) - \mu\sigma\mathrm{v}\left(\frac{\partial B_y}{\partial y}\right) \tag{15.130}$$

To obtain a similar equation in E_z and/or J_z, we differentiate the first equation of (15.128) partially with respect to y, and the second one (again partially) with respect to x, and get

$$\frac{\partial^2 E_z}{\partial y^2} = -\left(\frac{\partial^2 B_x}{\partial y \partial t}\right) + \mathrm{v}\left(\frac{\partial^2 B_x}{\partial y^2}\right),$$

and

$$-\left(\frac{\partial^2 B_z}{\partial x^2}\right) = -\left(\frac{\partial^2 B_y}{\partial x \partial t}\right) - \mathrm{v}\left(\frac{\partial^2 B_x}{\partial x^2}\right)$$

Subtracting the second equation from the first,

$$\begin{aligned}\frac{\partial^2 E_z}{\partial x^2} + \frac{\partial^2 E_z}{\partial y^2} &= -\left(\frac{\partial}{\partial t}\right)\left(\frac{\partial B_x}{\partial y} - \frac{\partial B_y}{\partial x}\right) + \mathrm{v}\left[\frac{\partial^2 B_x}{\partial y^2} + \left(\frac{\partial}{\partial x}\right)\left(-\frac{\partial B_y}{\partial y}\right)\right] \\ &= -\left(\frac{\partial}{\partial t}\right)(-\mu\sigma E_z) + \mathrm{v}\left(\frac{\partial}{\partial y}\right)\left(\frac{\partial B_x}{\partial y} - \frac{\partial B_y}{\partial x}\right) \\ &= +\mu\sigma\left(\frac{\partial E_z}{\partial t}\right) + \mathrm{v}\left(\frac{\partial}{\partial y}\right)(-\mu\sigma E_z)\end{aligned}$$

or

$$\left(\frac{\partial^2 E_z}{\partial x^2}\right) + \left(\frac{\partial^2 E_z}{\partial y^2}\right) = \mu\sigma\left(\frac{\partial E_z}{\partial t}\right) - \mu\sigma\mathrm{v}\left(\frac{\partial E_z}{\partial y}\right) \tag{15.131}$$

and hence for J_z,

$$\left(\frac{\partial^2 J_z}{\partial x^2}\right) + \left(\frac{\partial^2 J_z}{\partial y^2}\right) = \mu\sigma\left(\frac{\partial J_z}{\partial t}\right) - \mu\sigma\mathrm{v}\left(\frac{\partial J_z}{\partial y}\right) \tag{15.132}$$

Since the induced fields in the block are due to a travelling field source, the solutions for the magnetic field and the current density vectors would take the form:

$$\begin{aligned}\mathbf{B} &= \mathrm{Re}\left[\mathbf{i}_x \hat{B}_x(x) + \mathbf{i}_y \hat{B}_y(x)\right]\exp\left[j(\omega_S t - k_1 y)\right] \\ \mathbf{J} &= \mathrm{Re}\left[\mathbf{i}_z \hat{J}_z(x)\right]\exp\left[j(\omega_S t - k_1 y)\right]\end{aligned} \tag{15.133}$$

Substituting for B_x in Eq. (15.130),

$$\frac{d^2 \hat{B}_x}{dx^2} - \alpha_1^2 \hat{B}_x = 0 \tag{15.134}$$

where

$$\alpha_1 = k_1\sqrt{1 + j\left(\frac{\mu\sigma}{k_1^2}\right)(\omega_S + \mathrm{v}k_1)} = k_1\sqrt{1 + jS}$$

$$S = \left(\frac{\mu\sigma}{k_1^2}\right)(\omega_S + \mathrm{v}k_1) \tag{15.135}$$

The solution of Eq. (15.134) is

$$\hat{B}_x = \hat{B}_1 \exp(-\alpha_1 x) + \hat{B}_2 \exp(+\alpha_1 x) \tag{15.136}$$

where α_1 has been defined to have a positive real part. The x-dimension of the block is assumed to be quite large compared to the wavelength of the travelling current sheet, (i.e. $2\pi/k_1$), and hence for the first boundary condition:

1. As $x \to \infty$, $B_x \to 0$ $\quad \therefore \quad B_2 = 0$, and hence:

$$\hat{B}_x = \hat{B}_1 \exp(-\alpha_1 x) \tag{15.137}$$

2. The second boundary condition is: on $x = 0$, $\hat{B}_y = \mu_0 A$

where A will be determined by the amplitude of the imposed current sheet. We also assume that the air-gap between the applied current sheet and the surface of the conducting block is quite small, and the lower surface of the current sheet is bounded by a highly permeable region. Hence to apply this boundary condition, we use Eq. (15.129) from which we get

$$\hat{B}_y = \left(\frac{1}{jk_1}\right)\left(\frac{\partial \hat{B}_x}{\partial x}\right) = -\left(\frac{\alpha_1}{jk_1}\right)\hat{B}_1$$

$\therefore$ On $x = 0$, $\hat{B}_y = \mu_0 A$, and hence:

$$\hat{B}_1 = -\left(\frac{jk_1}{\alpha_1}\right)\mu_0 A$$

$$\therefore \quad \mathbf{B} = \mathrm{Re}\left[-\mathbf{i}_x\left(\frac{jk_1}{\alpha_1}\right)\mu_0 A + \mathbf{i}_y \mu_0 A\right] \exp(-\alpha_1 x) \exp[(j\omega_S t - k_1 y)] \tag{15.138}$$

The induced currents in the moving block are obtained by using Eq. (15.122a), i.e.

$$\mathbf{J} = \mathbf{i}_z J_z = \mathbf{i}_z \hat{J}_z \exp\left[j(\omega_S t - k_1 y)\right] = \mathbf{i}_z \left(\frac{1}{\mu_0}\right)\left(\frac{\partial B_y}{dx} - \frac{\partial B_x}{\partial y}\right)$$

$$= \mathbf{i}_z \left(-\alpha_1 A + \frac{k_1^2}{\alpha_1}\right) \exp(-\alpha_1 x) \exp[j(\omega_S t - k_1 y)]$$

$$= \mathbf{i}_z j \left(\frac{\mu\sigma}{\alpha_1}\right)(\omega_S + \mathrm{v}k_1)\, A \exp(-\alpha_1 x) \exp[j(\omega_S t - k_1 y)] \tag{15.139}$$

J is, in fact, the real part of the above expression.

If α_1 is expressed as a complex number of the form

$$\alpha_1 = C + jD = k_1 \sqrt{\left[1 + j\left(\frac{\mu\sigma}{k_1^2}\right)(\omega_S + \mathrm{v}k_1)\right]} = k_1 \sqrt{(1 + jS)}$$

then

$$C = \left(\frac{k_1}{\sqrt{2}}\right)\left(\sqrt{1 + S^2} + 1\right)^{1/2}, \qquad D = \left(\frac{k_1}{\sqrt{2}}\right)\left(\sqrt{1 + S^2} - 1\right)^{1/2} \tag{15.140}$$

This shows that the term $\exp(-\alpha_1 x)$ can be expressed as $\exp(-Cx) \cdot \exp(-Dx)$, which implies that these $\mathbf{B}_x$, $\mathbf{B}_y$, $\mathbf{J}_z$ vectors, apart from travelling in the $+y$-direction in the conducting block, they also diffuse (i.e. or disperse or travel with attenuation) in the $+x$-direction, which means that the wavelength of the diffusing pattern in the x-direction is $(2\pi/D)$ and the pattern is being attenuated by a damping factor which is $\exp(-Cx)$, where C and D are given by Eq. (15.140). It will be noticed that both C and D are functions of S, i.e. $(\mu\sigma/k_1^2)(\omega_S + \mathrm{v}k_1)$, and the magnetic diffusion time will be a function of the wavelength of the diffusion wave which is a function of $(\mu\sigma/k_1^2)$; whereas the rate of change with respect to time of B_x for an observer moving with the velocity v of the slab (or the block) is

$$\left(\frac{\partial}{\partial t} - \mathrm{v}\frac{\partial}{\partial y}\right) B_x \rightarrow j(\omega_S + \mathrm{v}k_1)\,\hat{B}_x \tag{15.141}$$

i.e. $(\omega_S + \mathrm{v}k_1)$ is the frequency of the magnetic flux density for an observer moving with the conducting block.

$$\therefore \qquad S \propto \frac{\text{magnetic diffusion time}}{\text{period of excitation in the frame of the moving medium}}$$

i.e.
$$S \propto \left(\frac{\mu\sigma}{k_1^2}\right)(\omega_S + \mathrm{v}k_1) \tag{15.142}$$

The phase velocity of the travelling wave is ω/k, and S is zero when $\mathrm{v} = \omega_S/k_1$ and *is in the direction of travelling current sheet* (*i.e. the +ve y-direction*). Under this condition, there is no interaction between the conducting block and the travelling current sheet, which means that no currents are induced and the magnetic flux density completely penetrates the medium. When $\mathrm{v} = 0$, the situation is similar to that of the skin effect described in Section 15.2. However we are, at present, considering the situation when the conducting block is moving in the direction opposite to that of the exciting (travelling) current sheet.

Now if the velocity of the current sheet is $= \omega_S/k_1$ in the $+y$-direction and the conducting block is moving in the $-y$-direction with the velocity v whose magnitude is same as that of the exciting current sheet, i.e. $\mathrm{v} = \omega_S/k_1$, then the equations for the magnetic flux density **B** (15.138) and the current density **J** (15.139) in the block reduce to the form

$$\mathbf{B} = \left[-\mathbf{i}_x \frac{jk_1}{\alpha} + \mathbf{i}_y\right] \mu_0 A \exp(-\alpha x) \exp[j(\omega_S t - k_1 y)]$$

$$\mathbf{J} = \mathbf{i}_z j\left(\frac{2\omega_S \mu\sigma}{\alpha}\right) A \exp(-\alpha x) \exp[j(\omega_S t - k_1 y)]$$

where

$$\alpha = k_1\left(1 + j\,\frac{2\omega_S\mu\sigma}{k_1^2}\right)^{1/2} \tag{15.143}$$

These expressions will be considered in the subsequent sections of this chapter where they will be used as the basis for the equivalence of different approaches of loss calculations in turbogenerator rotors.

The results of the analysis of the present system (with some modifications, i.e. the conducting block and the travelling current sheet both travelling in the same $+y$-direction) are also usable in other important practical applications, such as tracked transportation schemes with levitation. This is because the induced currents in the conducting block produce time average forces in the x- and y-directions. The force in the x-direction makes it possible to levitate the block on the magnetic field, and the force in the y-direction can be used for accelerating the propulsion of the slab in that direction.

Note further that when ω_S is zero, then S becomes proportional to v [Eq. (15.142)], and in this limit S is a magnetic Reynold's number based on the wavelength of the current sheet. This indicates that a large magnetic Reynold's number implies that the induced field effects are significant.

15.12 TRAVELLING CURRENT SHEET IMPOSED ON A STATIONARY CONDUCTING MEDIUM

The geometry of the present problem is same as that of the problem discussed in the previous Section 15.11, the only differences being that the conducting block is now stationary (as shown in Figure 15.23), and the velocity of the travelling current sheet is twice that of its velocity in the previous problem. The travelling current sheet is

$$= \text{Re}[\mathbf{i}_z A \exp\{j(\omega_2 t - k_2 y)\}$$

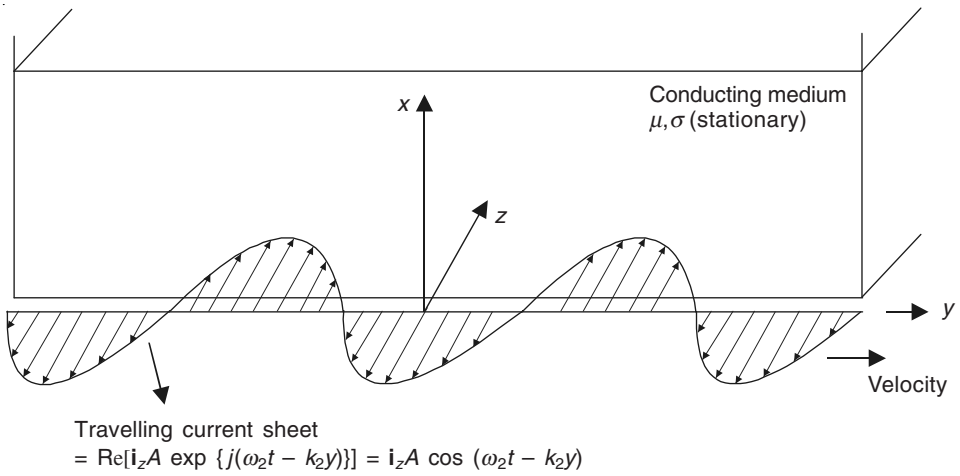


Figure 15.23 Stationary conducting medium, subjected to a travelling current sheet.

In the present problem, $\omega_2 = 2\omega_S$, of the previous problem and the wavelengths of both the current sheets are the same. Since the conducting block is not moving, i.e. v = 0, the operating equation which would be satisfied by the magnetic flux density **B** and the induced current density **J** would be of the form

$$\nabla^2 \mathbf{B} = \mu\sigma\left(\frac{\partial \mathbf{B}}{\partial t}\right) \tag{15.144}$$

As in the previous case, **J** will have only one component, i.e. J_z, and **B** will have the components B_x and B_y; and there will be no variations in the z-direction, i.e. $\partial/\partial z = 0$. Hence, we can derive by a process similar to that of the previous section (Section 15.11), that the equations satisfied by B_x, B_y, and J_z would be:

$$\frac{\partial^2 B_x}{\partial x^2} + \frac{\partial^2 B_x}{\partial y^2} = \mu\sigma\left(\frac{\partial B_x}{\partial t}\right) \tag{15.145a}$$

$$\frac{\partial^2 B_y}{\partial x^2} + \frac{\partial^2 B_y}{\partial y^2} = \mu\sigma\left(\frac{\partial B_y}{\partial t}\right) \tag{15.145b}$$

$$\frac{\partial^2 J_z}{\partial x^2} + \frac{\partial^2 J_z}{\partial y^2} = \mu\sigma\left(\frac{\partial J_z}{\partial t}\right) \tag{15.145c}$$

The solutions of **B** and **J** can be written in the form:

$$\mathbf{B} = \mathrm{Re}[\mathbf{i}_x \hat{B}_x(x) + \mathbf{i}_y \hat{B}_y(x)] \exp[j(\omega_2 t - k_2 y)]$$
$$\mathbf{J} = \mathrm{Re}[\mathbf{i}_z \hat{J}_z(x)] \exp[j(\omega_2 t - k_2 y)] \tag{15.146}$$

B also satisfies the equation

$$\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} = 0 \tag{15.147}$$

As before, substituting from Eq. (15.146) in Eq. (15.145), we get

$$\frac{\partial^2 \hat{B}_x}{\partial x^2} - \alpha_2^2 \hat{B}_x = 0$$

where

$$\alpha_2 = k_2 \left(1 + j\frac{\omega_2 \mu\sigma}{k_2^2}\right)^{1/2} = k_2 \left(1 + jS_2\right)^{1/2}$$
$$S_2 = \frac{\omega_2 \mu\sigma}{k_2^2} \tag{15.148}$$

The solution of the above equation is of the form

$$\hat{B}_x = \hat{B}_3 \exp(\alpha_2 x) + \hat{B}_4 \exp(-\alpha_2 x) \tag{15.149}$$

The relevant boundary conditions in this problem are:

1. On $x = 0$, $\hat{B}_y = \mu_0 A$, neglecting the air-gap dimensions; and
2. as $x \to \infty$, $\mathbf{B} \to 0$.

$$\therefore \qquad \hat{B}_3 = 0 \qquad \text{and} \qquad \hat{B}_x = \hat{B}_4 \exp(-\alpha_2 x) \tag{15.150}$$

From Eq. (15.147),

$$\hat{B}_y = \left(\frac{1}{jk_2}\right)\left(\frac{\partial B_x}{\partial x}\right) = -\left(\frac{\alpha_2}{jk_2}\right)\hat{B}_x$$

using Eq. (15.150).

And using B_y at $x = 0$ as $= \mu_0 A$, we get

$$\mathbf{B} = \left[-\mathbf{i}_x \left(\frac{jk_2}{\alpha_2}\right)\mu_0 A + \mathbf{i}_y \mu_0 A\right] \exp(-\alpha_2 x) \exp\left[j(\omega_2 t - k_2 y)\right] \tag{15.151}$$

$$\mathbf{J} = \mathbf{i}_z J_z = \mathbf{i}_z \hat{J}_z \exp[j(\omega_2 t - k_2 y)]$$
$$= \mathbf{i}_z \left[\left(\frac{\partial H_y}{\partial x}\right) - \left(\frac{\partial H_x}{\partial y}\right)\right] = \left(\frac{\mathbf{i}_z}{\mu}\right)\left[\left(\frac{\partial B_y}{\partial x}\right) - \left(\frac{\partial B_x}{\partial y}\right)\right]$$
$$= \mathbf{i}_z \left[-\alpha_2 A + \left(\frac{k_2^2}{\alpha_2}\right)A\right] \exp(-\alpha_2 x) \exp[j(\omega_2 t - k_2 y)]$$
$$= -\mathbf{i}_z j \left(\frac{\omega_2 \sigma\mu}{\alpha_2}\right) A \exp(-\alpha_2 x) \exp[j(\omega_2 t - k_2 y)] \tag{15.152}$$

In this case, $\omega_2 = 2\omega_S$, and $k_2 = k_1 = k$, then the expressions for the magnetic flux density and the eddy current density distributions in the conducting block become

$$\mathbf{B} = \left[-\mathbf{i}_x \left(\frac{jk}{\alpha_2} \right) + \mathbf{i}_y \right] \mu_0 A \exp(-\alpha_2 x) \exp[j(2\omega_S t - ky)] \tag{15.153a}$$

$$\mathbf{J} = -\mathbf{i}_z j \left(\frac{2\omega_S \sigma \mu}{\alpha_2} \right) A \exp(-\alpha_2 x) \exp[j(2\omega_S t - ky)] \tag{15.153b}$$

where

$$\alpha_2 = k \left[1 + j \left(\frac{2\omega_S \mu \sigma}{k_2} \right) \right]^{1/2} = k \left(1 + jS_2 \right)^{1/2} \tag{15.153c}$$

Note: For the current sheets of these two problems, $\omega_2 = 2\omega_S$ but $k_2 = k_1 = k$ (i.e. the same wave number), because the wavelength λ is same for both the current sheet waves.

It should be further noted that Eqs. (15.153) are same as Eqs. (15.143) except that these wave patterns are travelling at a speed $2\omega_S$ with respect to a stationary conducting block; whereas in the problem of Eqs. (15.143), the wave patterns of these field distributions were travelling in the forward (i.e. $+y$-direction) direction with a speed ω_S with respect to a fixed frame of reference, and the conducting block was also travelling at the same speed but in the opposite direction (i.e. $-y$-direction) with respect to the same fixed frame of reference. Furthermore, since α_2 of this equation is same as the α of Eq. (15.143), the diffusion (or the dispersion) of these field distributions in the x-direction (i.e. in the direction normal to the plane of the applied current sheet) in the conducting block, would be identical with that of the previous problem. So, since the variations with respect to x- and y-variables are same for both the sets of equations, the induced current and the magnetic field distributions are identical and hence the eddy current losses (as will be shown later) and also the forces come out to be the same.

15.13 STATIONARY, PULSATING, SINUSOIDALLY DISTRIBUTED CURRENT SHEET IMPOSED ON STATIONARY CONDUCTING BLOCK

The geometry of this problem is also the same as the two problems discussed in Sections 15.11 and 15.12, i.e. a semi-infinite conducting block of permeability μ and conductivity σ is subjected to a current sheet on the surface $x = 0$ (Figure 15.24). This current sheet is not a

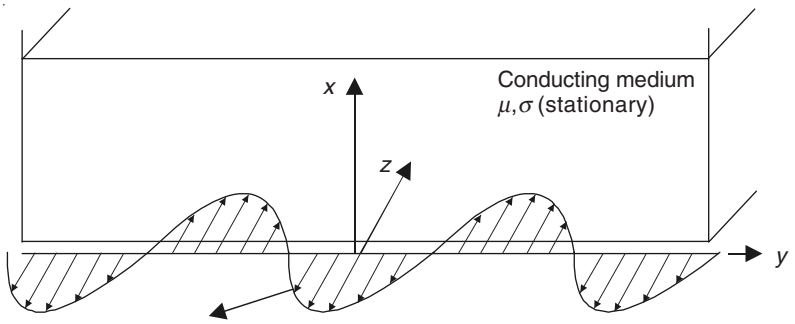


Figure 15.24 Stationary conducting block, subjected to a pulsating current sheet.

travelling current sheet as in the previous two problems, but is a stationary, pulsating (i.e. alternating and fixed spatially) sheet extending over the $x = 0$ surface. The wavelength of this current sheet is same as the wavelength of the travelling current sheets of those problems ($k_3 = k_2 = k_1 = k$).

Once again, **B** and **J** in the conducting block will be of similar form to what was derived in Section 15.10 or 15.11, and hence we shall not go into the derivation of the operating equations starting from the fundamental Maxwell's equations. These field vectors satisfy the vector diffusion equation of the form

$$\nabla^2 \mathbf{B} = \mu\sigma\left(\frac{\partial \mathbf{B}}{\partial t}\right) \tag{15.144}$$

as v = 0 in this case as well.

And the pulsating current sheet is = $\mathbf{i}_z A \cos k_3 y \exp(j\omega_3 t)$. (15.144′)

The frequency of this current sheet is same as that of Section 15.12, i.e. $\omega_3 = \omega_2 = 2\omega_S$.

As before, the current density vector **J** will have only the z-component J_z, and the magnetic flux density vector will have only B_x and B_y components, and these will have no variations in the z-direction [i.e. $(\partial/\partial z) = 0$]. The equations satisfied by these components would be same as the equations in (15.145) which we rewrite here for convenience:

$$\frac{\partial^2 B_x}{\partial x^2} + \frac{\partial^2 B_x}{\partial y^2} = \mu\sigma\left(\frac{\partial B_x}{\partial t}\right) \tag{15.145a}$$

$$\frac{\partial^2 B_y}{\partial x^2} + \frac{\partial^2 B_y}{\partial y^2} = \mu\sigma\left(\frac{\partial B_y}{\partial t}\right) \tag{15.145b}$$

$$\frac{\partial^2 J_z}{\partial x^2} + \frac{\partial^2 J_z}{\partial y^2} = \mu\sigma\left(\frac{\partial J_z}{\partial t}\right) \tag{15.145c}$$

The solutions for **B** and **J** for this problem would be of similar form to those of Section 15.12, but not the same, as the source is now a pulsating current sheet. Hence,

$$\begin{aligned}\mathbf{B} &= [\mathbf{i}_x \hat{B}_x(x) \sin k_3 y + \mathbf{i}_y \hat{B}_y(x) \cos k_3 y] \exp(j\omega_3 t) \\ \mathbf{J} &= [\mathbf{i}_z \hat{J}_x(x) \cos k_3 y] \exp(j\omega_3 t)\end{aligned} \tag{15.154}$$

B in this problem also satisfies the same zero divergence equation, i.e.

$$\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} = 0 \tag{15.147}$$

Hence substituting for B_x from Eq. (15.154) into the equation of (15.145), we get

$$\frac{d^2 \hat{B}_x}{\partial x^2} - \alpha_3^2 \hat{B}_x = 0$$

where

$$\begin{aligned}\alpha_3 &= k_3\left(1 + j\frac{\omega_3 \mu\sigma}{k_3^2}\right)^{1/2} = k_3\left(1 + jS_3\right)^{1/2} \\ S_3 &= \frac{\omega_3 \mu\sigma}{k_3^2}\end{aligned} \tag{15.155}$$

The expression for $\hat{B}_x$ will take the form

$$\hat{B}_x = \hat{B}_5 \exp(\alpha_3 x) + \hat{B}_6 \exp(-\alpha_3 x)$$

The boundary conditions at $x = 0$ and $x \to \infty$ are also same as before, i.e.

1. On $x = 0$, $\hat{B}_y = \mu_0 A$, neglecting the air-gap dimensions; and
2. as $x \to \infty$, $\mathbf{B} \to 0$.

$$\therefore \qquad \hat{B}_5 = 0 \qquad \text{and} \qquad \hat{B}_x = \hat{B}_6 \exp(-\alpha_3 x) \tag{15.156}$$

And from Eq. (15.147),

$$\hat{B}_y = \left(\frac{1}{jk_3}\right)\left(\frac{\partial \hat{B}_x}{\partial x}\right) = -\left(\frac{\alpha_3}{jk_3}\right)\hat{B}_x$$

and from the boundary condition (1), B_y at $x = 0$ as $= \mu_0 A$.

$$\therefore \qquad \mathbf{B} = \left[-\mathbf{i}_x \left(\frac{\mu_0 k_3}{\alpha_3}\right) A \sin k_3 y + \mathbf{i}_y \mu_0 A \cos k_3 y\right] \exp(-\alpha_3 x) \exp(j\omega_3 t)$$

and

$$\mathbf{J} = -\mathbf{i}_z j \left(\frac{\omega_3 \mu_0 \sigma}{\alpha_3}\right) A \cos k_3 y \exp(-\alpha_3 x) \exp(j\omega_3 t) \tag{15.157}$$

When $k_3 = k$, and $\omega_3 = 2\omega_S$, and since we are considering a conducting block, $\mu = \mu_0$, then $\alpha_3 = \alpha_2 = \alpha_1 = \alpha$.

So Eqs. (15.157) become

$$\mathbf{B} = \left[-\mathbf{i}_x \left(\frac{k}{\alpha}\right) \sin ky + \mathbf{i}_y \cos ky\right] \mu_0 A \exp(-\alpha x) \exp(j2\omega_S t)$$

$$\mathbf{J} = -\mathbf{i}_z j \left(\frac{2\omega_S \mu_0 \sigma}{\alpha}\right) A \cos ky \exp(-\alpha x) \exp(j2\omega_S t)$$

and

$$\alpha = k\left(1 + j\,\frac{2\omega_S \sigma \mu}{k^2}\right)^{1/2} = k\left(1 + jS_3\right)^{1/2} \tag{15.158}$$

Since the exp $(-\alpha x)$ term is same as in the previous problem (of Section 15.12), the dispersion of the field vectors in the x-direction (i.e. the radial direction in the turbogenerator rotor) is similar to those of the last problem.

15.14 LOSS CALCULATIONS IN THE CONDUCTING BLOCK

The three configurations, which we have considered in Sections 15.11 to 15.13, have given us the magnetic flux density **B** and the induced (eddy) current density **J** in the conducting block, under three different excitation systems. We shall now look into the losses in the block due to these excitations. The losses can be calculated by using the complex Poynting vector or the loss distribution using the formula $(\mathbf{J} \cdot \mathbf{J}^*/\sigma)$. Since in the practical problem (which occurs in the turbogenerator rotors and some other large ac rotating machines), the loss distribution patterns

are as important as the total losses, we shall, in the present analysis, use the (relatively more cumbersome) loss distribution formula, instead of the Poynting vector which calculates the total loss from the knowledge of the surface values of the field vectors.

So, for convenience, we rewrite the **B** and the **J** expressions from the three analyses, which we have completed just now. From Section 15.11, where the excitation is due to a travelling current sheet wave which is moving in the $+y$-direction at a velocity (ω_S/k) and the conducting block moving in the $-y$-direction with a velocity of same magnitude, we get from Eqs. (15.143).

$$\mathbf{B} = \left[-\mathbf{i}_x \left(\frac{jk}{\alpha} \right) + \mathbf{i}_y \right] \mu_0 A \exp[(-\alpha x) \exp(j\omega_S t - ky)]$$

$$\mathbf{J} = -\mathbf{i}_z j \left(\frac{2\omega_S \mu\sigma}{\alpha} \right) A \exp[(-\alpha x) \exp(j\omega_S t - ky)]$$

where

$$\alpha = k\left(1 + j\,\frac{2\omega_S \sigma\mu}{k^2}\right)^{1/2} = k(1 + jS_1)^{1/2} \tag{15.143}$$

From Section 15.12, where the excitation is due to a travelling current sheet wave moving with a velocity $(2\omega_S/k)$ in the $+y$-direction and a stationary conducting block, we get from Eqs. (15.153),

$$\mathbf{B} = \left[-\mathbf{i}_x \left(\frac{jk}{\alpha_2} \right) + \mathbf{i}_y \right] \mu_0 A \exp(-\alpha_2 x) \exp\, j(2\omega_S t - ky) \tag{15.153a}$$

$$\mathbf{J} = -\mathbf{i}_z j \left(\frac{2\omega_S \sigma\mu}{\alpha_2} \right) A \exp(-\alpha_2 x) \exp\, j(2\omega_S t - ky) \tag{15.153b}$$

where

$$\alpha_2 = k\left(1 + j\,\frac{2\omega_S \mu\sigma}{k^2}\right)^{1/2} = k(1 + jS_2)^{1/2} \tag{15.153c}$$

From Section 15.13, where the excitation is due to a pulsating current sheet which is alternating with an angular frequency $(2\omega_S)$ and has the same wavelength as $(1/k)$ as that of the travelling waves of the current sheets of the earlier two sections, and a stationary conducting block, we have the expressions for the field vectors as given in Eqs. (15.158) as

$$\mathbf{B} = \left[-\mathbf{i}_x \left(\frac{k}{\alpha} \right) \sin ky + \mathbf{i}_y \cos ky \right] \mu_0 A \exp(-\alpha x) \exp(j2\omega_S t)$$

$$\mathbf{J} = -\mathbf{i}_z j \left(\frac{2\omega_S \mu_0 \sigma}{\alpha} \right) A \cos ky \exp(-\alpha x) \exp(j2\omega_S t)$$

and

$$\alpha = k\left(1 + j\,\frac{2\omega_S \mu\sigma}{k^2}\right)^{1/2} = k(1 + jS_3)^{1/2} \tag{15.158}$$

Loss calculations. Since the time-averaging over one time-period of a sinusoidal wave pattern introduces merely a factor of ½ in both the travelling as well as the pulsating current sheet wave problems, we need to consider only the amplitude parts of these expressions for the induced

(or eddy) current distribution vectors. It is also seen that both the current distributions due to the travelling current sheets as obtained in the two sets of equations (15.143) and (15.153) have the same amplitudes, and hence it follows that the total losses as well as the loss distributions would be same in either of these two cases. So now, for comparison, we need to consider only two cases instead of the three, i.e. one due to a travelling current sheet and the other due to a pulsating current sheet. So we shall now use the suffix 'tr' (i.e. $\mathbf{J}_{tr}$) to indicate the induced current density obtained either from Eqs. (15.143) or (15.153), and 'pl' (i.e. $\mathbf{J}_{pl}$) to indicate the induced current density produced by the pulsating current sheet of Eqs. (15.158). So we have got the following two expressions:

$$\hat{\mathbf{J}}_{tr} = -\,\mathbf{i}_z\, j \left(\frac{2\omega_S \mu_0 \sigma}{\alpha} \right) A \exp(-\alpha x) \exp(-jky) \tag{15.159}$$

$$\hat{\mathbf{J}}_{pl} = -\,\mathbf{i}_z\, j \left(\frac{2\omega_S \mu_0 \sigma}{\alpha} \right) A \exp(-\alpha x) \cos ky \tag{15.160}$$

Since α is complex, we have

$$\alpha = k\left(1 + j\frac{2\omega_S \mu_0 \sigma}{k^2}\right)^{1/2} = k(1 + jS)^{1/2} = a + jb, \qquad \text{then}$$

$$a = \frac{1}{\sqrt{2}} \left[\left(k^4 + 4\omega_S^2 \mu_0^2 \sigma^2 \right)^{1/2} + k^2 \right]^{1/2}$$

$$b = \frac{1}{\sqrt{2}} \left[\left(k^4 + 4\omega_S^2 \mu_0^2 \sigma^2 \right)^{1/2} - k^2 \right]^{1/2}$$

$$\mathbf{J}_{tr} = -\,\mathbf{i}_z\, j \left[\frac{2\omega_S \mu_0 \sigma A}{\left(k^4 + 4\omega_S^2 \mu_0^2 \sigma^2 \right)^{1/2}} \right] \exp(-ax)(b + ja) \exp[-j(bx + ky)] \tag{15.161a}$$

$$\mathbf{J}_{pl} = -\,\mathbf{i}_z\, j \left[\frac{2\omega_S \mu_0 \sigma A}{\left(k^4 + 4\omega_S^2 \mu_0^2 \sigma^2 \right)^{1/2}} \right] \exp(-ax)(b + ja) [\exp\{(-jbx)\} \cos ky] \tag{15.161b}$$

$\therefore$ The loss density at any point is given by

$$\delta W_{tr} = \frac{\left(\hat{\mathbf{J}}_{tr} \cdot \hat{\mathbf{J}}_{tr}^* \right)}{\sigma} \tag{15.162a}$$

$$\delta W_{pl} = \frac{\left(\hat{\mathbf{J}}_{pl} \cdot \hat{\mathbf{J}}_{pl}^* \right)}{\sigma} \tag{15.162b}$$

where the asterisk (*) denotes the complex conjugate of the quantity. Let us also denote

$$\frac{2\omega_S \mu_0 \sigma A}{\left(k^4 + 4\omega_S^2 \mu_0^2 \sigma^2 \right)^{1/2}} = C \tag{15.163}$$

then

$$\delta W_{\text{tr}} = \left(\frac{C^2}{\sigma}\right)(a^2 + b^2)\exp(-2ax) \tag{15.164a}$$

$$\delta W_{\text{pl}} = \left(\frac{C^2}{\sigma}\right)(a^2 + b^2)\exp(-2ax)\cos^2 ky \tag{15.164b}$$

When we consider our model in these three problems to represent an idealized and simplified section of a turbogenerator rotor (under the specified operating conditions), the y-direction corresponds to the peripheral direction of the rotor, the z-direction corresponds to the axial direction of the machine, and the x-direction as the radial direction. So we shall now calculate the time-averaged loss over one pole-pitch length in the peripheral direction, i.e. in the y-direction from $y = 0$ to $y = \pi D_{\text{R}}/2$ [D_{R} being the rotor diameter of (say) a two-pole machine].

$$\therefore \qquad k = \frac{2\pi}{\lambda} = \frac{2\pi}{\pi D_{\text{R}}} = \frac{2}{D_{\text{R}}}$$

In the axial direction, we shall take unit length (as there is no z-variation, and the rotor is long enough to justify the neglect of the end effects and which have to be considered separately), and as the z-directed currents penetrate the rotor radially (i.e. the x-direction), we are justified in integrating over the radial distance from $x = 0$ (the rotor surface) to $x \to \infty$ (i.e. towards the shaft centre-line).

$$\bar{W}_{\text{tr}} = \frac{1}{2}\int_{x=0}^{x\to\infty}\int_{y=0}^{y=\pi D_{\text{R}}/2} \delta W_{\text{tr}}\, dx\, dy \cdot 1$$

$$= \left[\frac{C^2(a^2 + b^2)}{2\sigma}\right]\left(\frac{\pi D_{\text{R}}}{2}\right)\int_0^\infty \exp(-2ax)\, dx$$

$$= \left[\frac{C^2(a^2 + b^2)\,\pi D_{\text{R}}}{4\sigma}\right]\left(\frac{1}{-2a}\right)(0 - 1) = \frac{C^2(a^2 + b^2)\,\pi D_{\text{R}}}{8a\sigma} \tag{15.165}$$

$$\bar{W}_{\text{pl}} = \frac{1}{2}\int_{x=0}^{x\to\infty}\int_{y=0}^{y=\pi D_{\text{R}}/2} \delta W_{\text{pl}}\, dx\, dy \cdot 1$$

$$= \left[\frac{C^2(a^2 + b^2)}{2\sigma}\right]\int_0^{\pi D_{\text{R}}/2} \cos^2 ky\, dy \int_0^\infty \exp(-2ax)\, dx$$

$$= \left[\frac{C^2(a^2 + b^2)}{2\sigma}\right]\left(\frac{1}{2}\right)\left(\frac{\pi D_{\text{R}}}{2}\right)\left(\frac{1}{2a}\right) = \frac{C^2(a^2 + b^2)\,\pi D_{\text{R}}}{16a\sigma} \tag{15.166}$$

The bar "–" indicates the time-averaged value of the losses per time-period which introduces the factor of (½) in each of the above expressions. We find that:

$$\bar{W}_{\text{tr}} = \bar{W}_{\text{pl}} \left[\frac{\omega_S^2 \mu_0^2 \sigma \pi D_{\text{R}} A^2}{2a\left(k^4 + 4\omega_S^2 \mu_0^2 \sigma^2\right)^{1/2}} \right] \tag{15.167}$$

The above expression shows the equivalence relationship between the results due to these two (different) excitations imposed on the same system, and thus provides the mathematical and physical justifications for using the stationary current sheet approach to solve the problems with travelling current sheet excitations as well as with moving conducting media.

15.15 EDDY CURRENT LOSSES AND SATURATION EFFECTS IN IRON

The eddy current losses (or the induced current losses) calculated on the basis of the constant relative permeability in ferromagnetic media have been found to be too small when they are compared with the measured values. This is quite a complicated problem, which has been investigated by a number of workers, like, MacLean, McConnel, and Agarwal. When the iron is in the form of thin sheets, i.e. laminations in transformers and ac rotating machines, the domain structure becomes one of the most important factors. This aspect plays an important role in the anisotropic, cold-rolled, and grain-oriented (CRGO) steel laminations used in power transformers. The anomaly between the calculated eddy current losses (under assumed constant μ) and the measured losses was found to be (by Brailsford) ranging between 3 to 10 times. This loss was named as the 'anomalous eddy current loss', and the anomaly factor was found to be a function of the domain size, frequency and $B_{\text{max.}}$

Whilst the domain effects are important in thin sheets, in solid iron and steel the deciding factor is the shape of the B–H relationship. In the present discussion, we shall emphasize the underlying principle of how to deal with the saturation effects of B, rather than the detailed results. First we consider a semi-infinite slab of iron (Figure 15.25), on whose $x = 0$ surface an alternating magnetizing force $H_{\text{s}} \sin \omega t$ is applied (directed in the +ve y-direction). At this stage we make a drastic simplifying assumption regarding the B–H characteristic of the saturated material, i.e. that the B–H curve is a step-function as indicated in Figure 15.26. Hence under the steady-state conditions, the flux density then can have only two values, i.e. either $+B_{\text{s}}$ or $-B_{\text{s}}$.

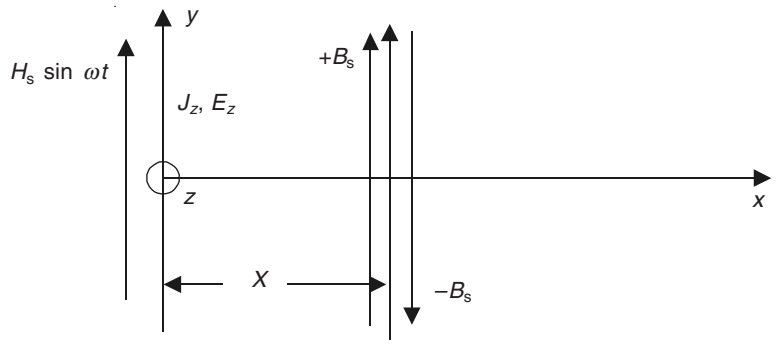


Figure 15.25 Penetration of flux in a semi-infinite iron block.

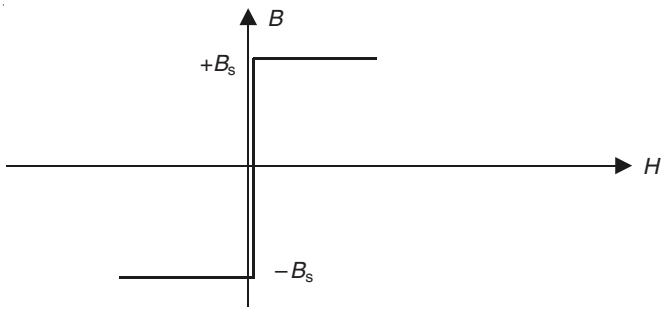


Figure 15.26 Step-function *B–H* characteristic of the material.

When a sinusoidally time-varying magnetizing field $H_s \sin \omega t$ is imposed on the surface of the block, then during the positive half-cycle, the influence of the positive magnetic field penetrates into the material. Thus a wave of magnetic flux moves into the medium, and at the wave-front there is a change from $+B_s$ to $-B_s$. And this wave is followed by one of opposite sign, and such a process keeps on repeating with the time-period of the alternating field.

In the semi-infinite block under consideration, at an instant of time t, let the distance of the wavefront from the surface of the material be X, and hence its velocity is (dX/dt). This transition surface where the flux density changes from B_s to $-B_s$ has been called the 'separating surface' by Agarwal in his analysis. Now the rate of change of flux (= Φ) per unit width in the z-direction is

$$\frac{d\Phi}{dt} = 2B_s \left(\frac{dX}{dt}\right) \tag{15.168}$$

Beyond this distance X, at this moment, there are no time-varying effects, because the wave has not penetrated so far. Hence there are no induced currents for $x > X$. Hence applying the Faraday's law (Maxwell's curl **E** equation) to a rectangular circuit, which encloses the wavefront, we have

$$E_z = -2B_s \left(\frac{dX}{dt}\right) \quad \text{and} \quad J_z = -2\sigma B_s \left(\frac{dX}{dt}\right) \tag{15.169}$$

$\therefore$ The total current per unit width in the y-direction is

$$I = -H_s \sin \omega t = \int_0^X J_z \cdot dx \cdot l = -2\sigma B_s X \left(\frac{dX}{dt}\right) \tag{15.170}$$

or

$$\sigma B_s \left(\frac{d(X)^2}{dt}\right) = H_s \sin \omega t$$

$\therefore$ Integrating with respect to time, we get

$$X^2 = \left(\frac{H_s}{\omega \sigma B_s}\right)(1 - \cos \omega t)$$

and hence

$$X = \left(\frac{2H_s}{\omega \sigma B_s}\right)^{1/2} \sin\left(\frac{\omega t}{2}\right) \tag{15.171}$$

where $t = 0$ is the instant of start of a positive half-cycle. The maximum depth of penetration occurs when $\sin(\omega t/2) = 1$ or $t = \pi/\omega$.

$$\therefore \qquad \delta_s = \left(\frac{2H_s}{\omega\sigma B_s}\right)^{1/2} = X_{max} \tag{15.172}$$

Comparing this with the skin-depth of the linear theory, i.e.

$$\delta_1 = d\sqrt{2} = \left(\frac{2}{\omega\mu_0\mu_r\sigma}\right)^{1/2} \tag{15.173}$$

it should be noted that in the nonlinear case, the permeability $\mu_0\mu_r$ has been replaced by the ratio B_s/H_s.

The velocity of the wavefront is [from Eq. (15.171)]:

$$\frac{dX}{dt} = \left(\frac{1}{2}\right)\omega\sqrt{\left(\frac{2H_s}{\omega\sigma B_s}\right)}\cos\left(\frac{\omega t}{2}\right) \tag{15.174}$$

It is also observed that the current density throughout the depth X is independent of the variable x.

$\therefore$ The instantaneous power loss per unit surface area is $= J_z^2 \dfrac{X}{\sigma}$.

$\therefore$ The energy lost per half-cycle $= W_1 = \dfrac{1}{\sigma}\displaystyle\int_{t=0}^{t=\pi/\omega} J_z^2\, X dt$

Substituting from Eqs. (15.169), (15.171), and (15.174),

$$W_1 = \left(\omega^2\sigma\delta_s^3 B_s^2\right)\int_{t=0}^{t=\pi/\omega}\cos^2\left(\frac{\omega t}{2}\right)\sin\left(\frac{\omega t}{2}\right)dt$$

$$= \left(\frac{2}{3}\right)\left(\omega^2\sigma\delta_s^3 B_s^2\right) \tag{15.175}$$

We now plot $E(t)$ [or $J(t)$], $X(t)$, and $H(t)$ as a function of t (Figure 15.27).

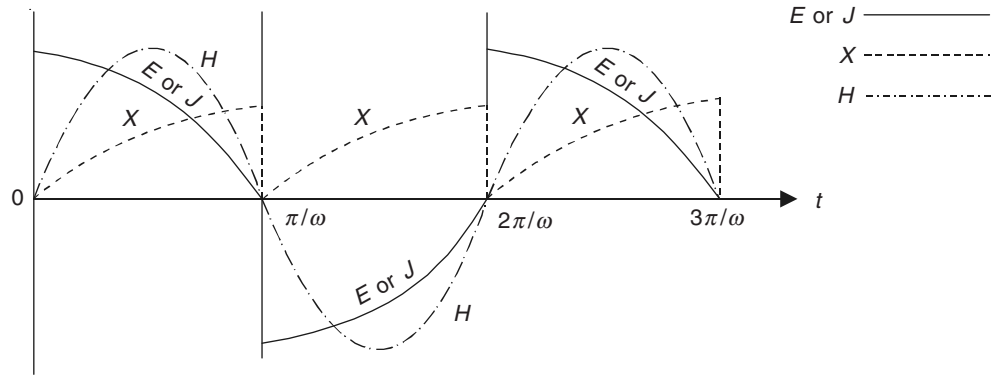


Figure 15.27 Plots of wave-shapes of E (or J), X and H in a thick saturated block of iron.

By a Fourier series analysis of H and E wave-shapes, the fundamental components come out to be

$$H_{1s} = -jH_s \qquad \text{and} \qquad E_{1s} = \left(\frac{8H_s}{3\pi\delta_s\sigma}\right)(\cos\ \omega t + 2\ \sin\ \omega t) \tag{15.176}$$

Comparing this with the loss calculated on the basis of constant μ it is seen that this loss is 70% higher than the linear B–H characteristic. This analysis holds for sheets whose half-thickness d (the thickness being $2d$) is $> \delta_s$. For a detailed analysis of the losses in thin sheets, interested readers are recommended to read the papers by Agarwal, and by Lim & Hammond.

15.16 PROXIMITY LOSS

The non-uniform distribution of current in an isolated conductor caused by the passage of an alternating current in the medium has been called the 'skin effect' due to the tendency of the current to concentrate near the surface layers of the conducting body. If a second conductor is present in the vicinity of the first conductor, both carrying alternating currents of same frequency, then the alternating current in the second conductor causes a further change in the current distribution in the first conductor. This phenomenon is known as the 'proximity effect'. Essentially both are electromagnetic induction phenomenon, and the distinction between them is, though clear, rather artificial. Since the proximity effect and the skin effect are really two facets of the same phenomenon, it is not always possible to analyze the two effects separately. Also the analysis of the combined proximity and skin effects is often quite complicated. For example, an analytical solution does exist for determining the proximity effects in a circular coaxial cable in which the inner and the outer conductors are carrying currents in the opposite directions (Figure 15.28). But the solution is quite involved, and is obtained as a function of complex Bessel's functions or Ber and Bei functions. However, in general, the influence of the

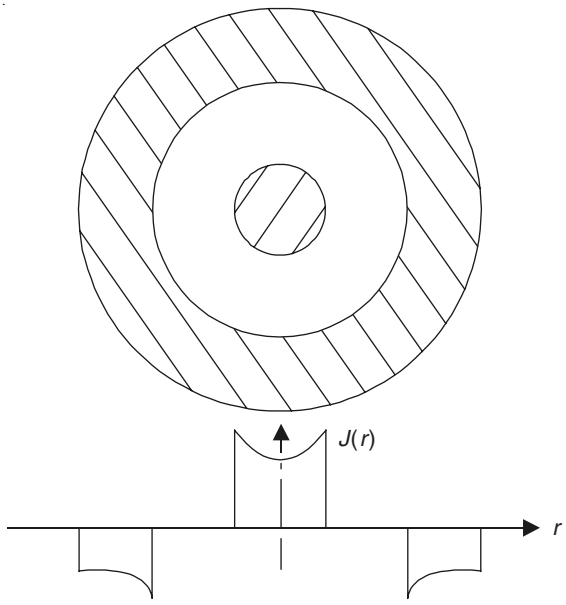


Figure 15.28 Proximity effect in a coaxial cable.

proximity effect is small compared with the skin effect. We shall, however, consider a highly idealized problem to illustrate this effect so as to gain some insight into this problem. We consider two conducting sheets of conductivity σ, permeability μ, and thickness b. The currents in the two sheets are equal in magnitude but opposite in direction, and their frequency is $\omega = 2\pi f$; and the sheets are so large that their finite size effects can be neglected. Since these sheets are large in size, we can assume them to be of infinite size, and hence the magnetic field on either side of each sheet is independent of the distance from the sheet (Figure 15.29). This follows from the problem discussed in the Chapter 7, Section 7.10.9.

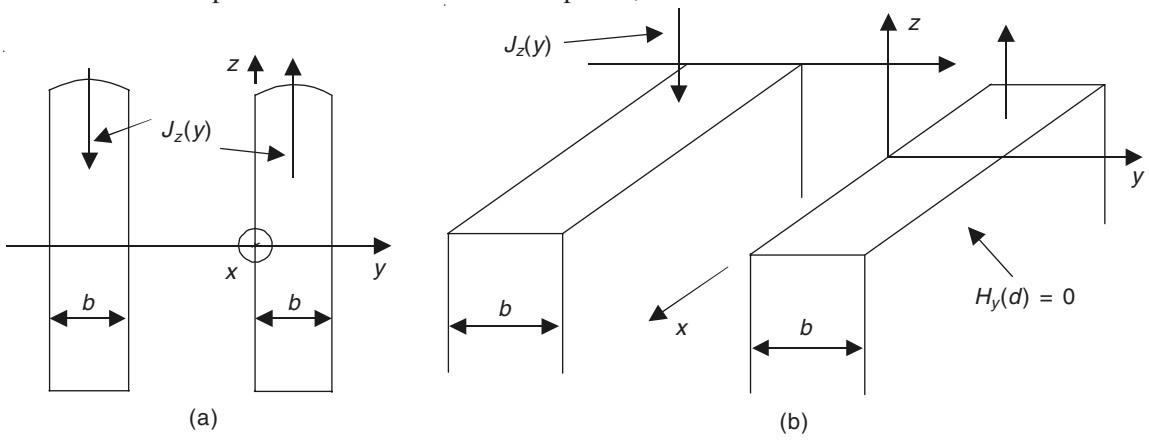


Figure 15.29 Two parallel strip conductors of thickness b carrying the same alternating current in opposite directions: (a) sectional view in the plane $x = 0$ and (b) isometric view.

Because of the symmetry of the problem, **J** and **H** would vary in the y-direction only (inside the sheets). Since the coordinate axes are as shown in Figure 15.29, the operational equation for **J** and **H** inside either sheet is obtained as

$$\frac{d^2 J_z}{dy^2} - j\omega\mu\sigma J_z = 0$$

and using the skin-depth notation $d^2 = (\omega\mu\sigma)^{-1}$, the equation for J_z becomes:

$$\frac{d^2 J_z}{dy^2} - \left(\frac{j}{d^2}\right) J_z = 0 \tag{15.177}$$

It should be noted that **J** would have only the z-component, and **H** only the x-component.

Its solution is of the form

$$J_z(y) = J_1 \exp\left[\frac{\sqrt{j}\, y}{d}\right] + J_2 \exp\left[\frac{-\sqrt{j}\, y}{d}\right] \tag{15.178}$$

Also, because of the symmetry of the problem, we need consider the current distribution in one of the two sheets. Let us consider the sheet with the current in the + z-direction. The **H** field in the sheet is given by the equation

$$j\omega\mu\sigma H_x = -\frac{dJ_z}{dy}$$

$$\therefore \qquad H_x(y) = -\left(\frac{1}{j\omega\mu\sigma}\right)\left(\frac{\sqrt{j}}{d}\right)\left\{J_1 \exp\left(\frac{\sqrt{j}\,y}{d}\right) - J_2 \exp\left(\frac{-\sqrt{j}\,y}{d}\right)\right\} \tag{15.179}$$

On $y = b$, $H_x(b) = 0$ [see Section 7.10.9, Eq. (7.68)].

$$\therefore \quad \text{This condition gives:} \quad J_2 = J_1 \exp\left(\frac{2\sqrt{j}\,b}{d}\right) \tag{15.180}$$

$$\therefore \qquad J_z(y) = J_1 \exp\left(\frac{\sqrt{j}\,b}{d}\right)\left[\exp\left\{\frac{\sqrt{j}\,(b-y)}{d}\right\} + \exp\left\{\frac{-\sqrt{j}\,(b-y)}{d}\right\}\right] \tag{15.181}$$

Expressing J_1 in terms of its value at $y = 0$, i.e. $J_z(0)$, we get

$$J_z(y) = J_z(0)\left[\frac{\exp\left\{\frac{\sqrt{j}\,(b-y)}{d}\right\} + \exp\left\{\frac{-\sqrt{j}\,(b-y)}{d}\right\}}{\exp\left\{\frac{\sqrt{j}\,b}{d}\right\} + \exp\left\{\frac{-\sqrt{j}\,b}{d}\right\}}\right] \tag{15.182}$$

Noting that $\sqrt{j} = \dfrac{1+j}{\sqrt{2}}$, we get

$$J_z(y) = J_z(0) \times \left[\frac{\cosh\left(\frac{b-y}{d\sqrt{2}}\right)\cos\left(\frac{b-y}{d\sqrt{2}}\right) + j\sinh\left(\frac{b-y}{d\sqrt{2}}\right)\sin\left(\frac{b-y}{d\sqrt{2}}\right)}{\cosh\left(\frac{b}{d\sqrt{2}}\right)\cos\left(\frac{b}{d\sqrt{2}}\right) + j\sinh\left(\frac{b}{d\sqrt{2}}\right)\sin\left(\frac{b}{d\sqrt{2}}\right)}\right]$$

$$\therefore \qquad \left|\frac{J_z(y)}{J_z(0)}\right| = \left[\frac{\cosh^2\left(\frac{b-y}{d\sqrt{2}}\right) - \sin^2\left(\frac{b-y}{d\sqrt{2}}\right)}{\cosh^2\left(\frac{b}{d\sqrt{2}}\right) - \sin^2\left(\frac{b}{d\sqrt{2}}\right)}\right]^{1/2} \tag{15.183}$$

This above ratio has been plotted as a function of (y/b), in Figure 15.30. It should be noted that if the second strip were not present, the current distribution would have been symmetrical about the line $y = b/2$. This analysis is based on the assumption of infinitely large sheets. If the finite dimensions of the sheets are taken into account, then the proximity effects become more pronounced when the sheets are nearer to each other. In general, the proximity effect tends to increase the resistance of the conductors and decrease the internal inductance.

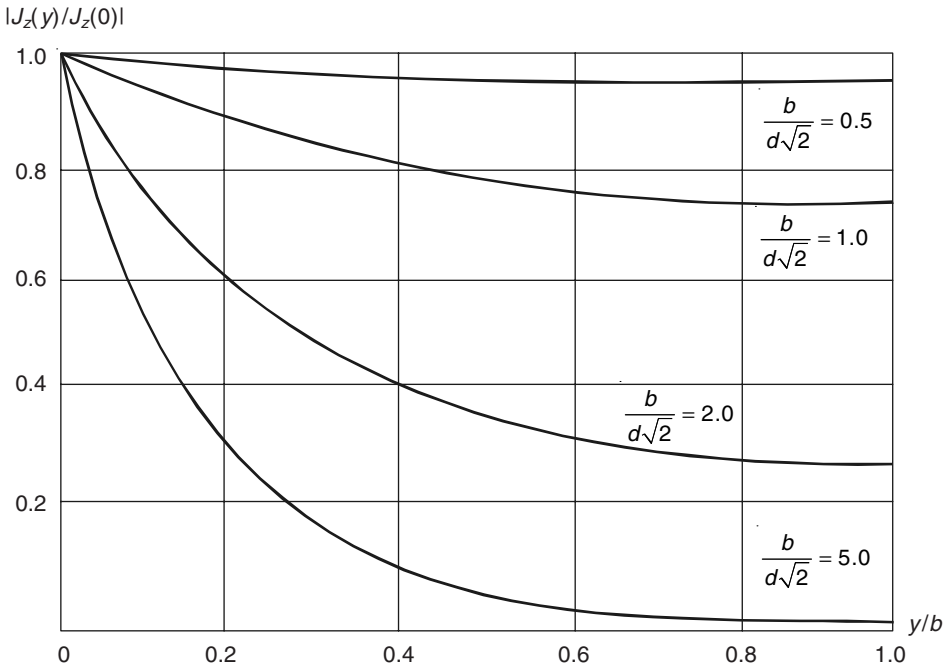


Figure 15.30 Current distribution in the strips of Figure 15.29.

15.17 UNILATERAL AND BILATERAL INDUCTION HEATING OF FLAT PLATES

Induction heating of flat metal plates is an industrial problem. There are other shapes of objects which are also heated by the method of induction, but most of possible shapes of objects can be reduced in theoretical and technical consideration to flat or cylindrical shape. The problem of heat treatment and induction heating has to be solved in the best possible way, which means optimizing the heating parameters for the electrothermal process, i.e. the three points for consideration are: (i) high heating rate, (ii) minimum electrical energy consumption, and (iii) requisite temperature distribution in the plate or medium. The frequency range used is usually 50 Hz to 3 MHz.

In industry, the objects are heated by induction either unilaterally or bilaterally, i.e. the suitably shaped heating inductor is applied to one side of the heated plate or object, or two inductors are applied to both the sides of the plate. The inductors are the source of the electromagnetic energy and can be represented (for the purpose of analysis) by a suitably directed current-sheet or its associated electric field (= **E**) or the magnetic field (= **H**). The induced field so generated in the heated plate is characterized by three vectors, i.e. the electric field **E** (or the induced current-density vector **J**), the associated magnetic field **H** (or **B**), and the Poynting vector (as a basis for the energy transfer from the inducing source).

In the present analysis, the object is a flat metal plate extending to infinity (for the purpose of analysis) both along its length and width (in the y- and z- directions respectively) and of finite thickness b (in the x-direction) The material characteristics of the plate are its electrical

conductivity σ $\left(\text{or resistivity } \rho = \frac{1}{\sigma}\right)$ and magnetic permability $\mu = \mu_0 \mu_r$ (μ_r being the relative permeability), both being assumed constants. Three different types of heating arrangements are considered here. First, the object is heated unilaterally (i.e. from one side only) by the inductor placed parallel to one of the faces. In the other two cases, the plate is heated bilaterally by means of two inductors placed on two sides of the plate. The two inductors are so positioned that the induced currents on the two faces of the plate are in the same direction and have the same phase. This is the case of bilateral heating with compatible currents. In the final arrangement, the current in one of the inductors is reversed so that the induced currents on the two faces of the plate are in opposite directions. In this case we have bilateral heating with inverse currents. The three arrangements are shown in Figure 15.31 (a), (b) and (c).

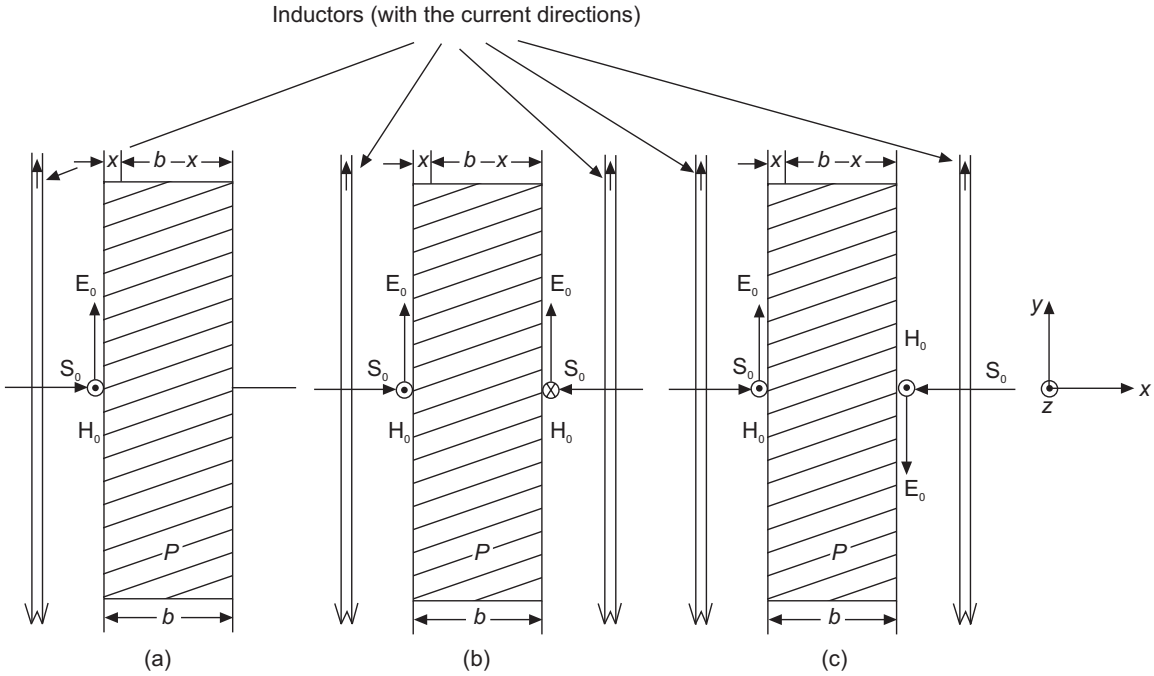


Figure 15.31 Induction heating systems of flat plate: (a) unilateral heating, (b) bilateral heating with compatible currents, and (c) bilateral heating with inverse currents.

The currents fed into the inductors are $I_0 e^{j\omega t}$, in the y-direction, so that the exciting **E** field on the surface of the plate is $\mathbf{i}_y E_o e^{j\omega t}$, $\omega = 2\pi f$, f being the frequency of the supply current within the range of the frequency used for induction heating (i.e. 50 Hz to 3 MHz). The associated magnetic field on the surface would be $\mathbf{i}_z H_0 e^{j\omega t}$. For the simplicity of analysis, the problem is reduced to a one-dimensional one so that the uniform current sheet extends to infinity in y- and z-directions, and the only variation of the fields is in the x-direction. So we shall now discuss the behaviour of these field vectors **E**, **H** and **S** (the Poynting vector = **E** × **H**) for these three systems, starting first with the unilateral heating system, because the other two arrangements can be considered as due to suitably modified superposition of the unilateral system applied to both the sides of the heated plate of finite thickness.

15.17.1 Unilateral Induction Heating of a Plate [Figure 15.31(a]

The characteristics of a heated flat metal plate of finite thickness would be known by determining the distribution of the electromagnetic field inside it, i.e. **E**, **J**, **H** and **S** in the plate. The geometry and the physics of the problem have been defined in Section 15.17. From Section 15.2, we know that when such an alternating magnetic or electric field is applied on the flat face of a semi-infinite metal block, the induced field in it is a travelling field, with its amplitude in the same direction as the inducing field, diffusing in the metal block in a direction normal to the plane on which the inducing field has been applied. By diffusion, we mean that the amplitude of the induced field gets attenuated as the field progresses inside [as seen in Eq. (15.12)]. In the present problem, the situation is somewhat different because now the conducting metal plate has finite thickness and it does not extend to infinity. So when the diffused field reaches the end surface of the plate (i.e. $x = b$), the field re-enters the air-space which has zero conductivity and the field in this region will become a plane uniform wave travelling with constant amplitude in this region and the fields on the two sides of this plane of discontinuity $x = b$ would satisfy the continuity conditions of **B** and **E** (or **H** and **J**). In this sense, this problem is more complex than that of the semi-infinite metal block as it is now a composite problem involving diffusion as well as uniform plane wave. (As such it should be studied after Chapter 17, though of course we have been briefly introduced to uniform plane waves in Chapter 12, Section 12.6.) So for a proper rigorous analysis of this problem, we start with Maxwell's equations though we could have written down the diffusion equation directly.

As mentioned earlier this is a simplified one-dimensional problem with variations in the x-direction only.

Hence $\frac{\partial}{\partial y} = 0$ and $\frac{\partial}{\partial z} = 0.$ The relevant Maxwell's equations are

$$\text{curl } \mathbf{E} = \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{15.4b}$$

$$\text{curl } \mathbf{H} = \nabla \times \mathbf{H} = \mathbf{J} \tag{15.4c}$$

and the constitutive relations are

$$\mathbf{B} = \mu\mathbf{H}, \qquad \mathbf{D} = \varepsilon\mathbf{E} \qquad \text{and} \qquad \mathbf{J} = \sigma\mathbf{E} \quad (\sigma = 1/\rho) \tag{15.4f}$$

The inductor current sheet is assumed to be in the y-direction, i.e.

$$\mathbf{J}_S = \mathbf{i}_y J_{So} e^{j\omega t} \tag{15.184a}$$

where

J_{So} = surface current density (per unit width)

ω = the angular frequency of the alternating current = $2\pi f$

Hence the inducing **E** field will be

$$\mathbf{E}_i = \mathbf{i}_y E_0 e^{j\omega t} \tag{15.184b}$$

and the associated magnetic field on the surface of the plate will be

$$\mathbf{H}_i = \mathbf{i}_z H_0 e^{j\omega t} \tag{15.184c}$$

Hence, Eqs. (15.4b) and (15.4c) reduce to

$$\mathbf{i}_x 0 + \mathbf{i}_y 0 + \mathbf{i}_z \frac{\partial E_y}{\partial x} = -\mathbf{i}_z \frac{\partial B_z}{\partial t}$$

and

$$\mathbf{i}_x 0 + \mathbf{i}_y \left(-\frac{\partial H_z}{\partial x} \right) + \mathbf{i}_z 0 = \mathbf{i}_y J_y$$

inside the plate.

Differentiating these equations w.r.t. x (which in this case is equivalent to taking curl again),

$$\frac{\partial^2 E_y}{\partial x^2} = -j\omega \frac{\partial B_z}{\partial x} = -j\omega\mu \frac{\partial H_z}{\partial x} = +j\omega\mu J_y = +j\omega\mu\sigma E_y$$

and

$$-\frac{\partial^2 H_z}{\partial x^2} = \sigma \frac{\partial E_y}{\partial x} = -j\omega\sigma B_z = -j\omega\mu\sigma H_z$$

from the constitutive relations (15.4f) and $\frac{\partial}{\partial t} = j\omega$

Now, $\omega\mu\sigma = \frac{1}{d^2}$, $d\sqrt{2}$ being the depth of penetration of the metal.

We substitute $j\omega\mu\sigma = k^2$ (or β^2)

$\therefore$ $$k = \frac{1+j}{d\sqrt{2}} \qquad (18.184d)$$

$\therefore$ The equations for E and H become:

$$\frac{\partial^2 E_y}{\partial x^2} + k^2 E_y = 0 \qquad \text{and} \qquad \frac{\partial^2 H_z}{\partial x^2} k^2 H_z = 0 \qquad (15.185)$$

In the air-space beyond the thickness of the plate (thickness = b and so the air-space is $x > b$), Eq. (15.4b) is replaced by

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} \qquad (15.4b')$$

and so the operating equation for **H** or **E** in this region will be

$$\frac{\partial^2 H_z}{\partial x^2} = j^2\omega^2\mu\varepsilon H_z \qquad \text{and} \qquad \frac{\partial^2 E_y}{\partial x^2} = j^2\omega^2\mu\varepsilon E_y$$

Substituting $-k_1^2 = j^2\omega^2\mu\varepsilon$, we get

$$\frac{\partial^2 E_y}{\partial x^2} + k_1^2 E_y = 0 \qquad \text{and} \qquad \frac{\partial^2 H_z}{\partial x^2} + k_1^2 H_z = 0 \qquad (15.186)$$

Since both **E** and **H** have one component only, we drop the suffices y and z and use the suffix 1 for the air-space to differentiate it from the plate.

So, the solutions for Eqs. (15.185) are

$$H = A \exp(-kx) + B \exp(kx)$$

and

$$\mathbf{E} = \frac{1}{\sigma}(\nabla \times \mathbf{H}) = -\frac{1}{\sigma}\frac{\partial H}{\partial x} \tag{15.187}$$
$$= \frac{k}{\sigma}\{A \exp(-kx) - B \exp(kx)\}$$

in the metal plate (i.e. $0 < x < b$).

In the air-space behind the metal plate (i.e. $x > b$)

$$\left.\begin{aligned} H_1 &= C \exp(-k_1 x) + D \exp(k_1 x) \qquad x > b \\ E_1 &= -\frac{1}{j\omega\varepsilon}\frac{\partial H_1}{\partial x} \\ &= \frac{k_1}{j\omega\varepsilon}\{C \exp(-k_1 x) - D \exp(k_1 x)\} \end{aligned}\right\} \tag{15.188}$$

Here A, B, C, D are the constants of integration to be determined from the boundary conditions and the interface continuity conditions:

(i) for $x \to \infty$, $\mathbf{H}_1$, $\mathbf{E}_1 \to 0$;
(ii) for $x = 0$, $H = H_0$;
(iii) for $x = b$, $\quad H_{b-} = H_{1_{b+}} \quad$ and $\quad E_{b-} = E_{1_{b+}}$

From the boundary condition (i), $D = 0$. It should be noted that the term containing D, in fact, represents the reflected wave which cannot exist in this case as the air-space extends to infinity. From (ii),

$$H_0 = A + B \tag{15.189}$$

From (iii),

$$A \exp(-kb) + B \exp(kb) = C \exp(-k_1 b) \tag{15.190a}$$

and

$$\frac{k}{\sigma}\{A \exp(-kb) - B \exp(kb)\} = \frac{k_1}{j\omega\varepsilon} C \exp(-k_1 b) \tag{15.190b}$$

From Eqs. (15.190a) and (15.190b),

$$\frac{A \exp(-kb) + B \exp(kb)}{A \exp(-kb) - B \exp(kb)} = \frac{k}{\sigma}\frac{j\omega\varepsilon}{k_1} \tag{15.191}$$

R.H.S of this equation

$$= \frac{j\omega\varepsilon k}{\sigma k_1} = \frac{j\omega\varepsilon}{\sigma}\frac{\sqrt{j\omega\mu\sigma}}{j\omega\sqrt{\mu\varepsilon}}$$
$$= \sqrt{j\omega}\sqrt{\frac{\varepsilon}{\sigma}} = \frac{1+j}{\sqrt{2}}\sqrt{2\pi f}\sqrt{\varepsilon\rho}$$

Since the frequency range of induction heating is 50 Hz to 3 MHz, and for metals $\varepsilon = \varepsilon_0 = 8.854 \times 10^{12}$ and ρ ranges from 1.75×10^{-6} Ω-cm for copper to approximately 140×10^{-6} Ω-cm for molten cost-iron,

the R.H.S. of Eq. (15.191) $\simeq 10^{-10} \simeq 0$

$$\therefore \qquad B = -A \exp(-2kb)$$

and hence

$$A = H_0 \frac{1}{1 - \exp(-2kb)}, \quad B = -H_0 \frac{\exp(-2kb)}{1 - \exp(-2kb)} \tag{15.192}$$

and from Eq. (15.190b) $C = 0$

Hence the intensity of the magnetic field behind the metal plate is zero, i.e. the magnetic field (and the associated electric field) does not penetrate through the metal plate for the range of industrial frequencies and the types of metals used in induction heating processes. The metal plate behaves like a magnetic screen.

By substituting for A and B in Eqs. (15.187), the field parameters inside the metal plate come out to be

$$H = H_0 \frac{\exp(-kx) - \exp(kx)\exp(-2kb)}{1 - \exp(-2kb)}$$

$$= H_0 \frac{\sinh\{k(b-x)\}}{\sinh(kb)} \tag{15.193a}$$

$$E = H_0 \frac{k}{\sigma} \frac{\cosh\{k(b-x)\}}{\sinh(kb)} \tag{15.193b}$$

$$J = H_0 k \frac{\cosh\{k(b-x)\}}{\sinh(kb)} \tag{15.193c}$$

J being the current density inside the plate.

The power density inside the plate would be

$$P = \frac{1}{2}\rho J^2$$

$$= \frac{1}{2} \frac{H_0^2 k^2}{\sigma} \frac{\cosh^2\{k(b-x)\}}{\sinh^2(kb)} \tag{15.193d}$$

For graphical representation of these quantities, we normalize them with respect to the surface values and take their moduli, remembering that k is complex, i.e. $= (1 + j)/(d\sqrt{2})$. So, we get

$$\frac{|H|}{|H_0|} = \sqrt{\left[\frac{\cosh\left\{\frac{2(b-x)}{d\sqrt{2}}\right\} - \cos\left\{\frac{2(b-x)}{d\sqrt{2}}\right\}}{\cosh\left(\frac{2b}{d\sqrt{2}}\right) - \cos\left(\frac{2b}{d\sqrt{2}}\right)}\right]} \tag{15.194a}$$

$$\frac{|E|}{|E_0|} = \frac{|J|}{|J_0|} \sqrt{\left[\frac{\cosh\left\{\frac{2(b-x)}{d\sqrt{2}}\right\} + \cos\left\{\frac{2(b-x)}{d\sqrt{2}}\right\}}{\cosh\left(\frac{2b}{d\sqrt{2}}\right) - \cos\left(\frac{2b}{d\sqrt{2}}\right)}\right]} \tag{15.194b}$$

and

$$\frac{|P|}{|P_0|} = \frac{\cosh\left\{\dfrac{2(b-x)}{d\sqrt{2}}\right\} + \cos\left\{\dfrac{2(b-x)}{d\sqrt{2}}\right\}}{\cosh\left(\dfrac{2b}{d\sqrt{2}}\right) - \cos\left(\dfrac{2b}{d\sqrt{2}}\right)} \tag{15.194c}$$

From the above expression, we can calculate total power per unit cross-sectional area along the total width b of the plate, i.e.

$$P_T = \int_0^b P\,dx = \frac{1}{\sigma}\left(\frac{H_0}{d\sqrt{2}}\right)^2 \int_0^b \frac{\cosh\left\{\dfrac{2(b-x)}{d\sqrt{2}}\right\} + \cos\left\{\dfrac{2(b-x)}{d\sqrt{2}}\right\}}{\cosh\left(\dfrac{2b}{d\sqrt{2}}\right) - \cos\left(\dfrac{2b}{d\sqrt{2}}\right)} dx$$

$$= \frac{H_0^2}{\sigma d\sqrt{2}} \frac{\sinh\left(\dfrac{2b}{d\sqrt{2}}\right) + \sin\left(\dfrac{2b}{d\sqrt{2}}\right)}{\cosh\left(\dfrac{2b}{d\sqrt{2}}\right) - \cos\left(\dfrac{2b}{d\sqrt{2}}\right)} \tag{15.195}$$

From this expression, it is seen that the total heating power of the plate is a function of H_0, the electrical parameters of the plate (i.e. σ), and also the ratio $\left(\dfrac{b}{d\sqrt{2}}\right)$, i.e. (thickness of the plate)/(depth of penetration of the eddy currents). The following quantity reduces to

$$\frac{\sinh\left(\dfrac{2b}{d\sqrt{2}}\right) + \sin\left(\dfrac{2b}{d\sqrt{2}}\right)}{\cosh\left(\dfrac{2b}{d\sqrt{2}}\right) - \cos\left(\dfrac{2b}{d\sqrt{2}}\right)} = \lambda = f\left\{\frac{b}{(d\sqrt{2})}\right\}. \tag{15.195a}$$

When the thickness of the plate b becomes infinite (i.e. much larger) relative to the depth of penetration $(= d\sqrt{2})$, then $\lambda \to 1$, and

$$\text{the heating power in the plate} = \frac{H_0^2}{\sigma\, d\sqrt{2}} \tag{15.196}$$

The variations of the heating power produced in the plate can be seen by studying the changes in λ as a function of the ratio of the thickness of the plate (= b) to its depth of penetration $(= d\sqrt{2})$. The function λ is near minimum when $\dfrac{b}{d\sqrt{2}} = \dfrac{\pi}{2}$, i.e. $\lambda \simeq 0.92$ and the heating power also reaches its minimum value which is 92% of the power defined by Eq. (15.196).

When we are considering the heating of a thick plate, we can let $b \to \infty$ (at least theoretically). The expressions for **H** and **E** simplify considerably, and the expressions become

$$\left.\begin{aligned} H &= H_0 \exp(-kx), \\ E &= \frac{k}{\sigma} H_0 \exp(-kx), \qquad J = kH_0 \exp(-kx) \end{aligned}\right\} \tag{15.197}$$

where k is now $= \dfrac{1}{d\sqrt{2}}$.

This is the case of the semi-infinite block which is discussed in detail in Section 15.2.

For the plate of finite thickness b, when $\left|\dfrac{H}{H_0}\right|$, $\left|\dfrac{E}{E_0}\right|$ and $\left|\dfrac{P}{P_0}\right|$ are plotted as functions of $\left|\dfrac{x}{b}\right|$ using $\dfrac{b}{d\sqrt{2}}$ as the parameter, it will be seen that H for all values of $\dfrac{b}{d\sqrt{2}}$, become zero when $x = b$, i.e. the plate acts as a screen preventing the leakage of magnetic field beyond its thickness. The magnetic field inside the plate is not uniform and has an exponential decay. The other quantities **E** and **P** decay near exponentially but have finite values at the other surface ($x = b$) of the plate.

15.17.2 Bilateral Induction Heating of the Flat Plate

The flat plate can be heated bilaterally in two different ways as shown in Figures 15.31(b) and (c), i.e. by means of currents induced on both the sides of a plate, in the same direction (= compatibly), or in opposite directions (= inversely). First we consider the case of compatible currents as shown in Figure 15.31(b). In that case the induced **E** fields on the two surfaces of the plate E_0' and E_0'' are equal in amplitude, phase and direction. But their associated magnetic fields H_0' and H_0'' will have opposite directions and equal amplitude and phase (why ?).

Inside the plate, the induced current will have the same direction as the electric field. At a point (say) x from the left surface of the plate [Figure 15.31b], the electric field intensities will be E' and E'', and the magnetic field intensities H' and H'', so that the resulting fields will be

$$H = H' - H'' \qquad \text{and} \qquad E = E' + E''$$

Each field can be evaluated as in the previous section (Section 15.17.1), and hence

$$H' = H_0 \frac{\sinh\{k(b-x)\}}{\sinh(kb)}, \qquad H'' = H_0 \frac{\sinh(kx)}{\sinh(kb)}$$

$\therefore$ The magnetic field at a point A (= x) will be

$$H = H_0 \frac{\sinh\{k(b-x)\} - \sinh(kx)}{\sinh(kb)}$$

$$= H_0 \frac{\sinh\{k(b/2 - x)\}}{\sinh(kb/2)} \tag{15.198}$$

Similarly the electric field

$$E = H_0 \frac{k}{\sigma} \frac{\cosh\{k(b/2 - x)\}}{\sinh(kb/2)} \tag{15.199}$$

and the current density

$$J = H_0 k \frac{\cosh\{k(b/2 - x)\}}{\sinh(kb/2)} \tag{15.200}$$

Heating power density inside the plate is also given by

$$P = \frac{1}{2\sigma} J^2 = \frac{1}{2\sigma} (H_0 k)^2 \frac{\cosh^2 \{k(b/2 - x)\}}{\sinh^2 (k\, b/2)} \tag{15.201}$$

As in Section (15.17.1), we normalize these quantities with respect to the surface values and take their moduli (k being complex as before),

$$\frac{|H|}{|H_0|} = \sqrt{\left[\frac{\cosh\left(\frac{b-2x}{d\sqrt{2}}\right) - \cos\left(\frac{b-2x}{d\sqrt{2}}\right)}{\cosh\left(\frac{b}{d\sqrt{2}}\right) - \cos\left(\frac{b}{d\sqrt{2}}\right)} \right]}, \tag{15.202a}$$

$$\frac{|E|}{|E_0|} = \frac{|J|}{|J_0|} = \sqrt{\left[\frac{\cosh\left(\frac{b-2x}{d\sqrt{2}}\right) + \cos\left(\frac{b-2x}{d\sqrt{2}}\right)}{\cosh\left(\frac{b}{d\sqrt{2}}\right) - \cos\left(\frac{b}{d\sqrt{2}}\right)} \right]}, \tag{15.202b}$$

and

$$\frac{|P|}{|P_0|} = \frac{\cosh\left(\frac{b-2x}{d\sqrt{2}}\right) + \cos\left(\frac{b-2x}{d\sqrt{2}}\right)}{\cosh\left(\frac{b}{d\sqrt{2}}\right) - \cos\left(\frac{b}{d\sqrt{2}}\right)} \tag{15.202c}$$

The fields have the maximum value on the two surfaces and decrease exponentially inside from both the surfaces, so that each component has a minimum at $x = b/2$. The magnetic inensity $H = 0$ at $x = b/2$.

The total heating power per unit cross-sectional area and along the whole width b is

$$\begin{aligned} P_T &= \int_0^b P dx = \frac{1}{2\sigma} (H_0 k)^2 \int_0^b \frac{\cosh^2 \left\{k\left(b/2 - x\right)\right\}}{\sinh^2 (kb/2)}\, dx \\ &= \frac{H_0^2}{4\sigma d^2} \int_0^b \frac{\cosh\left\{\frac{b-2x}{d\sqrt{2}}\right\} + \cos\left\{\frac{b-2x}{d\sqrt{2}}\right\}}{\cosh\left(\frac{b}{d\sqrt{2}}\right) - \cos\left(\frac{b}{d\sqrt{2}}\right)}\, dx \\ &= \frac{H_0^2}{2\sigma d\sqrt{2}} \frac{\sinh\left(\frac{b}{d\sqrt{2}}\right) + \sinh\left(\frac{b}{d\sqrt{2}}\right)}{\cosh\left(\frac{b}{d\sqrt{2}}\right) - \cos\left(\frac{b}{d\sqrt{2}}\right)} \end{aligned} \tag{15.203}$$

So, now the equivalent λ is:

$$\lambda = \frac{\sinh\left(\frac{b}{d\sqrt{2}}\right) + \sin\left(\frac{b}{d\sqrt{2}}\right)}{\cosh\left(\frac{b}{d\sqrt{2}}\right) - \cos\left(\frac{b}{d\sqrt{2}}\right)} \tag{15.204}$$

Next we consider a flat plate being heated inductively by two current sheets on two sides, the currents flowing in opposite directions [i.e. inversely, as shown in Figure [15.31(c)]. In this case the electric field intensities E_0' and E_0'' on the surfaces, will be of equal magnitude but of opposite directions whereas the magnetic field intensities H_0' and H_0'' will have the same magnitude and direction. So at any point inside the plate, the resultant field intensities will be

$$H = H' + H'' \quad \text{and} \quad E = E' - E''$$

As before,

$$H' = H_0 \frac{\sinh\{k(b-x)\}}{\sinh(kb)} \quad \text{and} \quad H'' = H_0 \frac{\sinh(kx)}{\sinh(kb)}$$

Hence the magnetic field at any point $A(= x)$ in the plate is

$$H = H_0 \frac{\sinh\{k(b-x)\} + \sinh(kx)}{\sinh(kb)}$$

$$= H_0 \frac{\cosh\{k(b/2-x)\}}{\cosh(kb/2)} \tag{15.205}$$

And similarly the electric field

$$E = H_0 \frac{k}{\sigma} \frac{\sinh\{k(b/2-x)\}}{\cosh(kb/2)} \tag{15.206}$$

and the current density

$$J = H_0 k \frac{\sinh\{k(b/2-x)\}}{\cosh(kb/2)} \tag{15.207}$$

In this arrangement, the heating power density in the plate

$$P = \frac{1}{2\sigma}(H_0 k)^2 \frac{\sinh^2\{k(b/2-x)\}}{\cosh^2(kb/2)} \tag{15.208}$$

As for the previous bilateral heating arrangement, normalizing these quantities with respect to their surface values and taking their moduli,

$$\frac{|H|}{|H_0|} = \sqrt{\left[\frac{\cosh\left(\frac{b-2x}{d\sqrt{2}}\right) + \cos\left(\frac{b-2x}{d\sqrt{2}}\right)}{\cosh\left(\frac{b}{d\sqrt{2}}\right) + \cos\left(\frac{b}{d\sqrt{2}}\right)}\right]} \tag{15.209a}$$

$$\frac{|E|}{|E_0|} = \frac{|J|}{|J_0|} = \sqrt{\left[\frac{\cosh\left(\frac{b-2x}{d\sqrt{2}}\right) - \cos\left(\frac{b-2x}{d\sqrt{2}}\right)}{\cosh\left(\frac{b}{d\sqrt{2}}\right) + \cos\left(\frac{b}{d\sqrt{2}}\right)}\right]} \tag{15.209b}$$

and

$$\frac{|P|}{|P_0|} = \frac{\cosh\left(\frac{b-2x}{d\sqrt{2}}\right) - \cos\left(\frac{b-2x}{d\sqrt{2}}\right)}{\cosh\left(\frac{b}{d\sqrt{2}}\right) + \cos\left(\frac{b}{d\sqrt{2}}\right)} \tag{15.209c}$$

As before, the fields have maximum values on the two surfaces of the plate, and decrease exponentially inside from both the surfaces, so that each component has a minimum at the centre $x = b/2$. The electric field intensity E and the current density J are zero at $x = b/2$.

Again, the total heating power per unit cross-sectional area and along the whole width b is

$$P_T = \int_0^b P dx = \frac{1}{2\sigma}(H_0 k)^2 \int_0^b \frac{\sinh^2\{k(b/2 - x)\}}{\cosh^2(kb/2)}\, dx$$

$$= \frac{H_0^2}{4\sigma d^2} \int_0^b \frac{\cosh\left\{\frac{b-2x}{d\sqrt{2}}\right\} - \cos\left\{\frac{b-2x}{d\sqrt{2}}\right\}}{\cosh\left(\frac{b}{d\sqrt{2}}\right) + \cos\left(\frac{b}{d\sqrt{2}}\right)}\, dx$$

$$= \frac{H_0^2}{2\sigma d\sqrt{2}} \frac{\sinh\left(\frac{b}{d\sqrt{2}}\right) - \sin\left(\frac{b}{d\sqrt{2}}\right)}{\cosh\left(\frac{b}{d\sqrt{2}}\right) + \cos\left(\frac{b}{d\sqrt{2}}\right)} \tag{15.210}$$

And the corresponding λ in this case is

$$\lambda = \frac{\sinh\left(\frac{b}{d\sqrt{2}}\right) - \sin\left(\frac{b}{d\sqrt{2}}\right)}{\cosh\left(\frac{b}{d\sqrt{2}}\right) + \cos\left(\frac{b}{d\sqrt{2}}\right)} \tag{15.211}$$

In all the three cases, i.e. (15.195a), (15.204) and (15.211), λ can be plotted as a function of $\left(\frac{b}{d\sqrt{2}}\right)$, to find the optimum value for the heating power in the plate, i.e. to maximize the eddy current losses in the plate. It should be noted that for these two types of bilateral induction heating, the magnetic field distribution in the plate with compatible currents [Eq. (15.198)], and the electric field or current density distribution [Eq. (15.206), Eq. (15.207)] for inverse currents are similar or nearly the same, both fields becoming zero at the central line (= $x = b/2$) of the plate. Similarly the electric field or current density distribution for compatible excitation current and the magnetic field distribution for inverse currents are same. However the loss distribution patterns for the two cases

are significantly different. For the inverse current system, the concentration of eddy currents nearer the two surfaces is much greater and the penetration into the depth is much less compared with the compatible current arrangement. So when the heating is to be restricted nearer the surfaces of the plate, the inverse current arrangement would be preferable.

Note: The inverse current system is the one which has been discussed is detail in Section 15.3, where the problem has been defined in terms of the applied magnetic fields on the two surfaces of the plate and the direction of this field is along the y-axis. In the present problem the magnetic field is directed along the z-axis.

15.18 EDDY CURRENTS AND SKIN EFFECT IN TUBULAR CONDUCTORS

It has been shown in Section 15.2 that when the depth of penetration is small compared with the curvature of the conducting surface, i.e. Eqs. (15.15) to (15.18) in Section 15.2, then the analysis for the plane surface can be applied to that of a tubular conductor. This is in general true for high frequency currents in tubular conductors. But if this is not true, then the problem has to be solved by using the cylindrical polar coordinate system. Before starting to solve the problem in this coordinate system, it is worth recapitulating the main points of the results for circular conductors (i.e. of circular cross-section) as derived in Section 15.2 in the equations mentioned above.

When the radius of a of the circular conductor, or the radial thickness (i.e. $a - b$) of a tubular conductor (of outer radius a and inner radius b) is much greater than the depth of penetration $d\sqrt{2}$ [$d = (\omega\mu\sigma)^{-1/2}$], then the ac resistance and the reactance of the conductor come out to be

$$R_{ac} = \omega L = \frac{1}{2\sqrt{2}\pi a\sigma d} \tag{15.212}$$

per unit axial length and the complete circumferential width (= $2\pi a$). Hence if a unit square is considered in the ϕ–z plane, then the ac resistance and reactance will be

$$R_{ac1} = \omega L_1 = \frac{1}{\sigma d\sqrt{2}} \tag{15.213}$$

It should be noted that the reactance (or the inductance) is due to the circumferential flux lying in the inside skin of the conductor where the alternating magnetic flux has penetrated starting from the outer surface.

Now, we consider the problem of the tabular conductor. In passing, it is mentioned that the problem of the isolated circular conductor has been solved in Section 15.5. The present problem does reduce to the problem of the circular conductor, when the inner radius b of the tube reduces to zero (i.e. $b \to 0$). Without going through all the intermediate steps, we can write down the equation for the current density $\mathbf{J}$ ($= \mathbf{i}_z J_z$ or $\mathbf{i}_z J$ to simplify the notations since only the z-component of the current in the axial direction of the conductor exists) as:

$$\mu\sigma \frac{d\mathbf{J}}{dt} = \nabla^2 \mathbf{J} \tag{15.214}$$

Since we are considering the steady state alternating current of angular frequency ω, i.e. $\mathbf{J}e^{j\omega t}$, the above equation further simplifies to

$$j\omega\mu\sigma J = \frac{d^2J}{dr^2} + \frac{1}{r}\frac{dJ}{dr} \tag{15.215}$$

as there is variation only in the r-direction and no variation in ϕ-direction due to symmetry about the axis of the conductors as well as in the z-direction (= the direction of current flow). Substituting $d^2 = \dfrac{1}{\omega\mu\sigma}$, the above equation is nothing but the modified Bessel's equation of zero order [cf. Eq. (4.44)]

i.e.
$$\frac{d^2J}{dr^2} + \frac{1}{r}\frac{dJ}{dr} + \left(\frac{j}{d}\right)^2 J = 0 \tag{15.215a}$$

and its solution is of the form

$$J = C\,I_0\left\{\left(\frac{j}{d^2}\right)^{1/2} r\right\} + D\,K_0\left\{\left(\frac{j}{d^2}\right)^{1/2} r\right\} \tag{15.216}$$

To evaluate C and D, the relevant boundary conditions are:

(i) On the outer boundary $r = a$, the current density J is equal to the surface current density J_0. Hence

$$J_0 = C\,I_0\left\{\sqrt{j}\,\frac{a}{d}\right\} + D\,K_0\left\{\sqrt{j}\,\frac{a}{d}\right\} \tag{15.217}$$

(ii) On the inner boundary $r = b$, there can be no magnetic field inside the cavity. This follows from the solution of the problem discussed in Section 7.10.3 [i.e. Eq. (7.58)]. There it is seen that the magnetic field in a circular cavity of a current-carrying circular conductor is uniform and is a function of the distance between the axis of the conductor and the axis of the circular cavity. Hence it follows that if the cavity is co-axial with the conductor, the magnetic field inside the cavity will be zero. Hence, at $r = b$, $\mathbf{B} = 0$

or
$$\frac{1}{\sigma}(\nabla \times \mathbf{J}) = -\frac{dB}{dt} = 0$$

or
$$\frac{dJ}{dr} = 0$$

$\therefore$
$$C\,I_1\left\{\sqrt{j}\,\frac{b}{d}\right\} + D\,K_1\left\{\sqrt{j}\,\frac{b}{d}\right\} = 0 \tag{15.218}$$

$\therefore$
$$C_1 = \frac{K_1\left\{\sqrt{j}\,\dfrac{b}{d}\right\} J_0}{J_0\left\{\sqrt{j}\,\dfrac{a}{d}\right\} K_1\left\{\sqrt{j}\,\dfrac{b}{d}\right\} + I_1\left\{\sqrt{j}\,\dfrac{b}{d}\right\} K_0\left\{\sqrt{j}\,\dfrac{a}{d}\right\}} \tag{15.219}$$

$$D_1 = \frac{I_1\left\{\sqrt{j}\,\frac{b}{d}\right\}}{I_0\left\{\sqrt{j}\,\frac{a}{d}\right\}K_1\left\{\sqrt{j}\,\frac{b}{d}\right\} + I_1\left\{\sqrt{j}\,\frac{b}{d}\right\}K_0\left\{\sqrt{j}\,\frac{a}{d}\right\}} \tag{15.220}$$

15.18.1 Skin Effect in Solid Cylindrical Conductors

If the conductor is a solid cylinder instead of a tube, then $b \to 0$ and $D = 0$ because $K_0\left\{\sqrt{j}\,\frac{r}{d}\right\} \to \infty$ as $r \to 0$. Hence the current distribution in the conductor is obtained as:

$$J = \frac{I_0\left\{\sqrt{j}\,\frac{r}{d}\right\}}{I_0\left\{\sqrt{j}\,\frac{a}{d}\right\}}J_0 \tag{15.221}$$

Since these modified Bessel's functions have complex arguments, they can be expressed as Kelvin's ber, bei, ker and kei functions (for the purpose of computations).

Thus,

$$I_0\left\{\sqrt{j}\,x\right\} = \text{ber}_0\,x + j\,\text{bei}_0\,x \tag{15.222a}$$

and

$$K_0\left\{\sqrt{j}\,x\right\} = \text{ker}_0\,x + j\,\text{kei}_0\,x \tag{15.222b}$$

This problem has been solved in detail in Section 15.5, where the Bessel functions with complex arguments have been expressed in polar form, and hence any further discussion of the above solution is not repeated here.

15.18.2 Internal Self-inductance of Cylindrical and Tubular Conductors due to Skin Effect While Carrying Alternating Currents

The analysis of the flux distribution in a cylindrical conductor (as well as in a tubular conductor with radial thickness greater than the "depth of penetration") carrying an alternating current has been discussed in Sections 15.2 and 15.18, and it is obvious that the magnetic flux loops are in the circumferential direction and inside the outer surface of the cylindrical as well as the tubular conductor. Hence the inductance caused by this alternating flux (as well as the effective ac resistance) can be considered as the "internal self-inductance" of the conductor (and also the internal resistance).

The ac inductance of the cylindrical (as well as tubular) wire as derived in Eq. (15.213) of Section 15.18 [deduced from Eq. (15.212) or Eq. (15.17) of Section 15.2] is effectively the inductance per unit axial length and the unit circumferential width, and can be denoted as L_{i1} and R_{i1}. These will now be expressed as:

$$L_{i1} = \frac{1}{\omega\sigma d\sqrt{2}} = \frac{R_{i1}}{\omega} \tag{15.223}$$

It should be noted that at high frequencies L_{i1} is negligible, compared with the external inductance of the wire (i.e. the inductance L being $L >> L_i$).

To evaluate the total L_i and R_i in practical problems such as cylindrical conductors in parallel wire transmission line problems, it is preferable to use energy expressions instead of using the flux linkage concept. The reason is that the presence of a second conductor in the vicinity of the first conductor produces a proximity effect which destroys the axial symmetry of the magnetic flux about the axis of each conductor. On the other hand the total current in the skin of the cylindrical conductor can be evaluated, with comparative ease. The expressions to be used are:

$$R_i I^2 = R_{i1} \oint J_s^2 ds \quad \text{and} \quad L_i I^2 = L_{i1} \oint J_s^2 ds \tag{15.224}$$

where J_s is the current across the unit arc length in the cylindrical skin of the conductor. The above expressions can be used for each of the two conductors, whatever may be the radius of each conductor. If the conductors are of equal diameter, the total inductance and resistance will be twice that obtained from the above formula. When the conductors are of equal diameter, the total inductance and resistance will be twice that obtained from the above formula. When the conductors are of unequal radii, R_i and L_i have to be evaluated separately by taking into account the relevant limits of integration. For such two-dimensional problems, the solution is most conveniently obtained by using conjugate functions and the method has been discussed in detail in Sections 4.4–4.4.3 and Section 17.18. A brief description of the important aspects of the results follows.

Since the equipotentials and the lines of force (or current flow lines) are orthogonal. We use the complex potential function W, and write

$$W = U + jV \qquad \text{and} \qquad W = f(z),\ z = x + jy \tag{15.225}$$

$$U = U(x, y), \qquad V = V(x, y) \tag{15.225a}$$

so that

$$U + jV = f(x + jy) \tag{15.226}$$

and

$$V - jU = jf(x + jy) = F(x + jy) \tag{15.226b}$$

$\therefore$ Differentiating W with respect to x and y respectively

$$\frac{\partial W}{\partial x} = \frac{\partial U}{\partial x} + j\frac{\partial V}{\partial x} = f'(z)\frac{\partial z}{\partial x} = f'(z) \tag{15.227a}$$

$$\frac{\partial W}{\partial y} = \frac{\partial U}{\partial y} + j\frac{\partial V}{\partial y} = f'(z)\frac{\partial z}{\partial y} = jf'(z) \tag{15.227b}$$

and hence,

$$\frac{\partial V}{\partial x} = -\frac{\partial U}{\partial y} \quad \text{and} \quad \frac{\partial V}{\partial y} = \frac{\partial U}{\partial x} \tag{15.228}$$

Thus, the two families of curves $U(x, y)$ = constant and $V(x, y)$ = constant intersect each other orthogonally. In such a case either set can be chosen to represent equipotentials in which case the chosen function becomes the potential function. The other set which is orthogonal to this then represents the lines of force (i.e. also called the stream function).

$$\therefore \quad \frac{dW}{dz} = \frac{\partial V}{\partial y} + j\frac{\partial V}{\partial x} = \frac{\partial U}{\partial x} - j\frac{\partial U}{\partial y} \left(\text{from } z = \frac{dW + jdV}{dx + jdy} \right) \tag{15.229}$$

Regardless of whether U or V is the potential function, the absolute value of $\frac{dW}{dz}$ at any point gives the magnitude of the electric field intensity at that point. From Eq. (15.229), it follows that

$$\left|\frac{dW}{dz}\right| = \frac{\partial U}{\partial n} = \frac{\partial V}{\partial s} \quad \text{or} \quad \left|\frac{dW}{dz}\right| = \frac{\partial V}{dn} = -\frac{\partial U}{\partial s} \tag{15.230}$$

(where dn is an element of length in the direction of maximum increase of potential, and ds the element of length obtained by rotating dn through $\pi/2$ radians in the counterclockwise sense), according as U or V is the potential function.

For example, if V is the potential function, then the flux through any section of an equipotential surface between the curves U_1 and U_2 is given by

$$\text{Flux} = -\varepsilon \int_{U_1}^{U_2} \frac{\partial V}{\partial n}\, ds = \varepsilon \int_{U_1}^{U_2} \frac{\partial U}{\partial s}\, ds = \varepsilon (U_2 - U_1) \tag{15.231}$$

Thus the capacitance of two closed equipotential surfaces at V_1 and V_2 can be expressed as

$$C = \frac{|Q|}{|V_2 - V_1|} = \frac{\varepsilon [U]}{|V_1 - V_2|} \tag{15.231}$$

where the increment in U is expressed as $[U]$. The inductance can then the evaluated using the relationship.

$$LC = \mu\varepsilon \tag{15.232}$$

To evaluate the internal inductance due to eddy current skin effect, the integral in Eq. (15.224) has to be evaluated. From Section 17.18, Eq. (17.236) the tangential component of **B** on the $U = U_1$ cylinder is $\frac{\partial V}{\partial s}$. The mmf applied to a surface element gives $\mu J_S = B$.

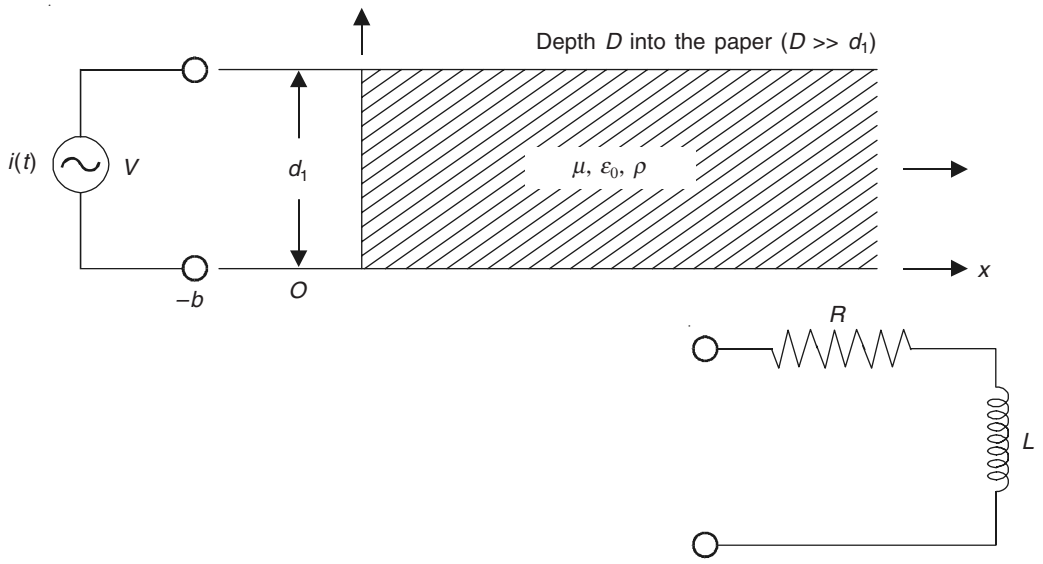
Hence,

$$\oint J_S^2 ds = \frac{1}{\mu^2} \oint \left[\frac{\partial V}{\partial s}\right]^2 ds = \frac{1}{\mu^2} \oint \frac{\partial V}{\partial s} dV \tag{15.233}$$

$\frac{\partial V}{\partial s}$ is known (for a given configuration) from the conjugate function relationship stated in Eq. (15.230), and hence the above integral can be suitably evaluated.

PROBLEMS

15.1 A pair of perfectly conducting plates holds a conducting block of rectangular cross-section, as shown in the figure below. The metal block and the plates extend a long way to the right. A current excitation $i(t) = \text{Re}\{I \exp(j\omega t)\}$ is applied uniformly to the plates along their left edge. Find the magnetic flux density in the region between the plates and the current density in the block.



Find the equivalent reactance as seen at the current source. Using the equivalent circuit shown here, find the values of L and R.

Hint: Solve as a one-dimensional problem in x dimension.

Ans.:
$$B_z = -\text{Re}\,\frac{\mu\hat{I}}{D}\exp\left(\frac{-x}{d\sqrt{2}}\right)\exp j\left(\omega t - \frac{x}{d\sqrt{2}}\right)$$

$$J_y = -\text{Re}\,\frac{\hat{I}}{D}\,\frac{1+j}{d\sqrt{2}}\exp\left(\frac{x}{d\sqrt{2}}\right)\exp j\left(\omega t - \frac{x}{d\sqrt{2}}\right)$$

$$R = \frac{d_1}{D}\,\frac{\rho}{d\sqrt{2}},\quad \omega L = \omega\left(\frac{\mu_0 b d_1}{D} + \frac{d_1}{D}\,\frac{\rho}{\omega d\sqrt{2}}\right),$$

$$\text{where } d = \left(\frac{\rho}{\omega\mu}\right)^{1/2}.$$

15.2 What will be the magnetic flux density and the current density in the above block if it extends over a length l in the x-direction?

Ans.: $$B_z = \frac{\mu \hat{I}}{D} \frac{\sinh \dfrac{(x-l)\sqrt{j}}{d}}{\sinh \dfrac{l\sqrt{j}}{d}}$$

$$J_y = \frac{\hat{I}}{D} \frac{\sqrt{j}}{d} \frac{\cosh (x-l) \dfrac{\sqrt{j}}{d}}{\sinh l\sqrt{\dfrac{j}{d}}}$$

Note: exp ($j\omega t$) is implicit in these expressions.

15.3 A copper conductor of strip form has a length and breadth which are both much greater than its thickness $2b$. A coordinate system is taken with its centre at the origin at the centre of all three dimensions of the strip, the axes of x, y and z being in the direction of the thickness, the breadth and the length respectively. The strip carries a current in the z-direction, the density being represented by a phasor J. Neglecting edge effects, prove that (by solving the equation for H or B),

$$H = H_y = J_0 d \left(\frac{1-j}{\sqrt{2}} \right) \sinh \frac{(1+j)\, x}{d\sqrt{2}}$$

where J_0 is the value of the current at the centre, and d is the skin-depth.

15.4 An alternating current flows longitudinally in a copper conductor of thickness $2b$ and resistivity ρ, the length and the width of the conductor being great compared with b. Prove that the ratio of the maximum to the minimum current density is

$$\left[\frac{1}{2} \left(\cosh \frac{b\sqrt{2}}{d} + \cos \frac{b\sqrt{2}}{d} \right) \right]^{1/2}$$

where $d = \left(\dfrac{\rho}{\mu_0 \omega} \right)^{1/2}$.

15.5 From the analysis of the current distribution in a semi-infinite conducting block, at radio frequencies, prove that the ratio of ac resistance to the zero-frequency resistance (i.e. dc resistance or R_{dc}) of conductors of any shape of cross-section is equal to the inverse ratio of the areas. Derive this expression for (i) a conductor of circular cross-section of radius a, (ii) a rectangular bar $a \times b$.

Hint: Use the one-dimensional skin-depth d.

15.6 An iron plate is bounded by the parallel planes $x = \pm b$. The plate extends to $+\infty$ in the z-direction and is wide enough in the $\pm y$-directions so that the edge effects can be ignored (which simplifies it to a one-dimensional problem). Wire is wound uniformly round the plate such that the layers of wire are parallel to the y-axis. An alternating current is sent through the wire, thus producing a magnetizing intensity $\mathbf{i}_z H_0 \cos \omega t$ on

the surfaces of the plate (i.e. **H** has only z-component on the two surfaces). Show that the **H** field inside the plate at a distance x from its centre is given by:

$$\mathbf{H} = \mathbf{i}_z H_0 \sqrt{\frac{\cosh 2mx + \cos 2mx}{\cosh 2mb + \cos 2mb}} \cdot \cos(\omega t + \beta)$$

$$\tan \beta = \frac{-\sinh m(b+x) \sin m(b-x) - \sinh m(b-x) \sin m(b+x)}{\cosh m(b+x) \cos m(b-x) + \cosh m(b-x) \cos m(b+x)}$$

where $m^2 = \dfrac{\omega\mu\sigma}{2} = \dfrac{1}{2d^2}$, i.e. $m = \dfrac{1}{d\sqrt{2}}$, d being the depth of penetration of iron of permeability $\mu(= \mu_0\mu_r)$ and conductivity σ. Discuss the limiting cases of mb small and mb large.

15.7 When there are time-varying currents in a conducting medium, the magnetic vector potential **A** satisfies the equation

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t}$$

Show, hence, that **A** satisfies the equation $\nabla^2 \mathbf{A} = \mu\sigma \dfrac{\partial \mathbf{A}}{\partial t}$ (a)

Show, now, the solution of **A** is of the form given by Eq. (13.7.2) of Problem 13.7, i.e.

$$\mathbf{A} = \nabla \times \mathbf{W} = \nabla \times (\mathbf{u} W_1 + \mathbf{u} \times \nabla W_2)$$

is such that **W**, W_1 and W_2 all satisfy the above equation (a). Also show that now W_1 and W_2 both contribute to the **B** field.

16 Charge Relaxation

16.1 INTRODUCTION

The relaxation of charge (in good or slightly conducting media) is the mechanism by which the motion has an effect on the electric field distributions in electric field systems. We shall discuss this phenomenon in this chapter by considering a number of examples. Initially we shall consider systems in which the conducting medium is at rest. In such problems, the relaxation time is of fundamental importance for determining the volume and the surface charge densities, which result from the initial conditions and excitations. Subsequently we shall look at the effect of steady motion on the relaxation of the free charges, and study the effects of the 'electric Reynold's number', based on the material velocity.

The field equations of relevance for the charge relaxation are

$$\text{curl } \mathbf{E} = \nabla \times \mathbf{E} = 0 \tag{16.1}$$

$$\text{div } \mathbf{D} = \nabla \cdot \mathbf{D} = \rho_{\text{fc}} \tag{16.2}$$

$$\text{div } \mathbf{J} + \frac{\partial \rho_{\text{fc}}}{\partial t} = \nabla \cdot \mathbf{J} + \frac{\partial \rho_{\text{fc}}}{\partial t} = 0 \tag{16.3}$$

Thus we consider two of the Maxwell's Eqs. (16.1) and (16.2). Equation (16.1) is the restricted form of the Faraday's law because here we are considering the electric field of *free charges* with which there is *no* associated time-varying magnetic field. The third equation (16.3) is the continuity equation (based on the conservation of charge) which is a consequence of the Maxwell's equations. The relevant constituent relations are

$$\mathbf{D} = \varepsilon \mathbf{E} \tag{16.4}$$

and we restrict ourselves to the situations in which the conduction of free charge can be accounted for by a constitutive law of the form

$$\mathbf{J} = \sigma \mathbf{E} + \rho_{\text{fc}} \mathbf{v} \tag{16.5}$$

in which we have combined the 'Ohm's law' with the field transformation arising out of motion. This law describes the conduction process in a wide range of solids, liquids and gases, but does not apply for all general situations.

As we analyze and study the subject of charge relaxation, we shall notice that this subject has some significant similarity with the magnetic diffusion, which we studied in Chapter 15. We now start with the equation of continuity [i.e. Eq. (16.3)] and the constituent relation given by Eq. (16.5)

$$\nabla \cdot (\sigma \mathbf{E} + \rho_{\text{fc}} \mathbf{v}) + \frac{\partial \rho_{\text{fc}}}{\partial t} = 0 \tag{16.6}$$

Expressing ρ_{fc} in terms of $\mathbf{E}$ by using Eqs. (16.2) and (16.4), we get

$$\nabla \cdot (\sigma \mathbf{E}) + \nabla \cdot [\mathbf{v} \nabla \cdot (\varepsilon \mathbf{E})] + \left(\frac{\partial}{\partial t} \right) [\nabla \cdot (\varepsilon \mathbf{E})] = 0 \tag{16.7}$$

When the velocity $\mathbf{v}$ is given, Eq. (16.7) involves only the electric field intensity $\mathbf{E}$. Also, it is a scalar equation and hence, in general, it will not uniquely define the three components of $\mathbf{E}$. However from Eq. (16.1), we can define a scalar potential ϕ such that

$$\mathbf{E} = - \text{ grad } \phi = - \nabla \phi \tag{16.8}$$

Then Eq. (16.7) becomes

$$\nabla \cdot (\sigma \nabla \phi) + \nabla \cdot [\mathbf{v} \nabla \cdot (\varepsilon \nabla \phi)] = - \left(\frac{\partial}{\partial t} \right) [\nabla \cdot (\varepsilon \nabla \phi)] \tag{16.9}$$

Physically this equation is an account of the conservation of the free charge in the system. The first term represents the flow of free charges into a small volume due to conduction. The second term is a consequence of the convection of the medium giving rise to the transport of the free charges into a given region. The term on the right-hand-side of the equation is the rate of increase of the local free-charge density. In the subsequent sections, we shall consider the effect of motion on the fields (i.e. when there is motion, the velocity is defined). However there are models (in problems of electromechanical couplings) in which the velocity $\mathbf{v}$ is not known till the fields are known.

16.2 CHARGE RELAXATION AS AN ELECTRICAL TRANSIENT

Initially we shall consider some problems dealing with the charge relaxation in stationary systems (i.e. when there is no motion, $\mathbf{v} = 0$). Also, we start with simpler systems in which the conductivity and the permittivity of each medium are uniform. Hence the changes in σ and ε occur only at the interface surfaces. It should also be appreciated that unless there is a free-charge source in the medium, there will be no steady-state volume free-charge density.

16.2.1 Charge Density in a Conducting Medium (or More Generally, Media with Uniform Properties)

Since $\mathbf{v} = 0$, the system equations are

$$\nabla \cdot \mathbf{J} + \frac{\partial \rho_{\text{fc}}}{\partial t} = 0 \tag{16.3}$$

$$\nabla \cdot \mathbf{D} = \rho_{\text{fc}} \tag{16.2}$$

$$\mathbf{D} = \varepsilon \mathbf{E} \tag{16.4}$$

$$\mathbf{J} = \sigma \mathbf{E} \tag{16.10}$$

Combining these four equations, we get the operational equation in ρ_{fc} as

$$\frac{\partial \rho_{fc}}{\partial t} + \left(\frac{\sigma}{\varepsilon}\right)\rho_{fc} = 0 \tag{16.11}$$

which can also be obtained from Eq. (16.9) or Eq. (16.7) by putting $\mathbf{v} = 0$ and substituting for $\mathbf{E}$ or ϕ in terms of the charge density $\rho_{fc.}$ A general solution of Eq. (16.11) is

$$\rho_{fc}(x, y, z, t) = \rho_0(x, y, z, 0) \exp\left(-\frac{t}{\tau}\right) \tag{16.12}$$

where

$$\tau = \frac{\varepsilon}{\sigma} \tag{16.13}$$

Equations (16.12) and (16.13) imply that given an initial charge density ρ_0 at the instant $t = 0$, the free-charge density at each point in space decays to zero exponentially with the 'relaxation time τ'. It should be understood that Eq. (16.11) has been derived without using Eq. (16.1), and hence these conclusions do not depend on the field equations being quasi-static. However in any physical situation, the uniformly conducting medium is of finite extent, and the conservation of charge requires that those charges which are initially distributed throughout the volume *relax to the surfaces* that bound the volume. We give below the relaxation time for some typical conductors and insulators (Table 16.1).

Table 16.1 Relaxation time for conductors and insulators

Material	σ (mho/metre)	ε	τ (seconds)
Silver	6.17×10^7	ε_0	1.43×10^{-19}
Copper	5.80×10^7	ε_0	1.52×10^{-19}
Aluminium	3.72×10^7	ε_0	2.38×10^{-19}
Mercury	1.06×10^6	ε_0	8.35×10^{-18}
Sea water	4	$80\varepsilon_0$	1.77×10^{-4}
Water	4×10^{-6}	$80\varepsilon_0$	6.40×10^{-4}
CCl_4	4×10^{-16}	$2.24\varepsilon_0$	4.95×10^4

For the conductors, the relaxation time is extremely small, and there are insulators for which the relaxation time can be measured in minutes and hours.

16.2.1.1 Example 1

A sphere of radius R_i is made up of uniformly conducting material, and is placed concentrically in another perfectly conducting spherical shell of radius R_o. The annular space between the spheres is filled with an isotropic insulating material. A free-charge density ρ_0 is distributed uniformly throughout a spherical region of radius $R(R < R_i)$ at the centre of the conducting sphere, at the instant of time $t = 0$.

$$\therefore \qquad \rho_{fc \text{ at } t=0} = \rho_0 = \frac{Q}{(4/3)\pi R^3} \tag{16.14}$$

where Q is the total charge in the central sphere of radius R.

∴ The transient charge density as given by Eq. (16.12) is

$$\rho_{fc} = \rho_0 \exp\left(-\frac{t}{\tau}\right) \qquad \text{for } r < R$$
$$= 0 \qquad \text{for } r > R \tag{16.15}$$

where $\tau = \varepsilon/\sigma$, with ε and σ being the permittivity and the conductivity respectively of the medium.

The problem has radial symmetry, i.e. the only variation is along r. Hence, using Eq. (16.2) in the spherical coordinate system, we obtain the equation for the electric field intensity E_r as, i.e.

$$\left(\frac{1}{r^2}\right)\frac{d}{dr}\left(r^2 E_r\right) = \left(\frac{\rho_0}{\varepsilon}\right)\exp\left(-\frac{t}{\tau}\right)$$

$$\therefore \qquad E_r = \frac{Q}{4\varepsilon\pi R^2}\left(\frac{r}{R}\right)\exp\left(-\frac{t}{\tau}\right) \qquad \text{for } r < R \tag{16.16}$$

The total charge within the radius $r = R$ is $Q \exp\left(-\frac{t}{\tau}\right)$, and by Gauss,

$$E_r = \left(\frac{Q}{4\varepsilon\pi r^2}\right)\exp\left(-\frac{t}{\tau}\right) \qquad \text{for } R < r < R_i \tag{16.17}$$

Since the total charge on the sphere must be conserved, the electric field outside the conducting sphere (in the insulating medium) is

$$E_r = \frac{Q}{4\varepsilon_0\pi r^2} \qquad \text{for } R_i < r < R_o \tag{16.18}$$

Hence from these electric field intensities, we can find the amount of free charge on the interface

$$\rho_S = D_n \underset{r \to R_i + \text{Eq. (16.18)}}{\text{(approaching from the insulating annulus)}} - D_n \underset{r \to R_i - \text{Eq. (16.17)}}{\text{(approaching from the conducting sphere)}}$$

$$= \frac{\varepsilon_0 Q}{4\varepsilon_0\pi R_i^2} - \left(\frac{\varepsilon Q}{4\varepsilon\pi R_i^2}\right)\exp\left(-\frac{t}{\tau}\right) = \frac{Q}{4\pi R_i^2}\left[1 - \exp\left(-\frac{t}{\tau}\right)\right] \tag{16.19}$$

The surface charge on the outer shell ($r = R_o$) is however constant. Equation (16.19) shows that the initial free-charge density relaxes to the outer surface of the conducting sphere. In the steady-state condition, there is *no* electric field within the sphere, as the field is shielded out of the sphere by the surface charge at $r = R_i$.

The current density is

$$J_r = \sigma E_r = \left(\frac{\sigma Q}{4\varepsilon\pi R^2}\right)\left[\left(\frac{r}{R}\right)\exp\left(-\frac{t}{\tau}\right)\right] \qquad \text{for } r < R$$

and

$$= \left(\frac{\sigma Q}{4\varepsilon\pi r^2}\right)\exp\left(-\frac{t}{\tau}\right) \qquad \text{for } R < r < R_i \tag{16.20}$$

It is this current, which accounts for the conduction of free charge to the surface of the sphere.

It should be noted that there is a conduction current in the region $R < r < R_i$ even though there is no free-charge density in this part of the conducting sphere during the transient.

16.2.1.2 *Example 2: Relaxation time of space charge in a partly filled tank*

It is generally believed that when the conductivity of the medium is constant, the relaxation time is independent of the size and the shape of the container in which the charged medium (being liquid or fluid) is enclosed. Asano in his studies of the relaxation of petroleum products found that when an open surface of charged medium exists within a conducting container, the relaxation time is a function of the shape. He also found that the ratio of the new relaxation time to the conventional relaxation time has a maximum value of 2, which is independent of the different permittivities of the two media. We give below the simplified one-dimensional analysis of Asano's problem (as derived by him). When charge is generated or introduced in dielectric liquids like hydrocarbons, the charge creates an electric field, which can cause an ignition hazard, and hence extensive studies have been undertaken in this field. In hydrocarbons, the charge relaxation process obeys the exponential charge relaxation law when the conductivity is high, and the hyperbolic law when σ is relatively low. For the former, a constant conductivity, and for the latter a constant mobility are assumed. The hyperbolic law applies when σ for the liquid is less than 1 pS/m. For the present problem under discussion, a constant σ and an exponential charge decay will hold.

A simplified model of a tank, which partly contains a hydrocarbon (charge) liquid, is shown in Figure 16.1. The model contains two regions: a liquid phase, and a vapour phase.

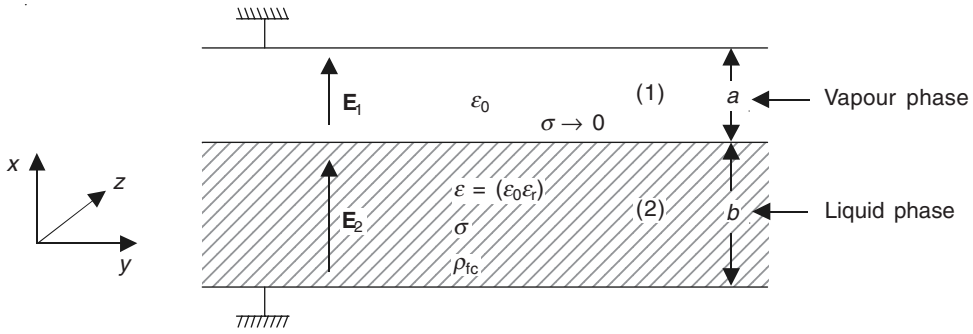


Figure 16.1 Simplified one-dimensional model of a tank containing charged liquid.

Let the charge density in the liquid be ρ_{fc}, conductivity σ, and permittivity $\varepsilon (= \varepsilon_0 \varepsilon_r)$. The free space parameters are ε_0 and $\sigma \to 0$. We assume that there is no surface charge on the interface between the liquid and the vapour.

The same three equations hold for this problem as well, i.e.

$$\nabla \times \mathbf{E} = 0 \tag{16.1}$$

$$\nabla \cdot \mathbf{D} = \rho_{fc} \tag{16.2}$$

$$\nabla \cdot \mathbf{J} = -\frac{\partial \rho_{fc}}{\partial t} \tag{16.3}$$

and the constituent relations are

$$\mathbf{J} = \sigma\mathbf{E} \qquad \text{and} \qquad \mathbf{D} = \varepsilon\mathbf{E} \tag{16.21}$$

From Eq. (16.1), it follows

$$\mathbf{E} = -\nabla\phi \tag{16.22}$$

where ϕ is a scalar potential.

$\therefore$ For the hydrocarbon, from Eqs. (16.22) and (16.2), we get

$$\nabla^2\phi = -\frac{\rho_{fc}}{\varepsilon} \tag{16.23}$$

and for the vapour phase

$$\nabla^2\phi = 0 \qquad \text{i.e. Laplace's equation} \tag{16.24}$$

For the region (1), solving the one-dimensional Laplace's equation

$$E_1 = \frac{d\phi_1}{dx} = -C_1$$

and integrating the above again

$$\phi_1 = C_1x + C_2 \tag{16.25}$$

where C_1 and C_2 are the constants of integration.

For the liquid region with the space charge,

$$E_2 = \frac{\rho_{fc}x}{\varepsilon} - D_1$$

$$\phi_2 = -\frac{\rho_{fc}x^2}{2\varepsilon} + D_1x + D_2 \tag{16.26}$$

where D_1 and D_2 are constants of integration.

Boundary conditions. 1. Since both the bottom and the top electrodes are grounded, they are at zero potential.

$$\therefore \qquad D_2 = 0 \qquad \text{and} \qquad C_2 = -C_1(a + b)$$

Hence the potentials for each region are

$$\phi_1 = C_1(x - a - b)$$

$$\phi_2 = -\frac{\rho_{fc}x^2}{2\varepsilon} + D_1x$$

2. On the interface $x = b$, both the potentials must be same, and as there is no surface charge on this plane, the normal component of the electric flux density must be continuous.

$$\therefore \qquad -C_1a = -\frac{\rho_{fc}b^2}{2\varepsilon} + D_1b \qquad \text{and} \qquad -\varepsilon_0C_1 = \varepsilon\left(\frac{\rho_{fc}b}{\varepsilon}\right) - D_1$$

$$\therefore \qquad C_1 = \frac{-\left(\dfrac{\rho_{fc}b}{2\varepsilon}\right)}{\dfrac{a}{b} + \dfrac{\varepsilon_0}{\varepsilon}}$$

and

$$D_1 = \frac{\left(\frac{\rho_{fc} b}{2\varepsilon}\right)\left(\frac{\varepsilon_0}{\varepsilon} + \frac{2a}{b}\right)}{\frac{a}{b} + \frac{\varepsilon_0}{\varepsilon}}$$

Hence the potentials in each region come out to be

$$\phi_1 = \frac{\frac{\rho_{fc} b}{2\varepsilon}(a + b - x)}{\frac{a}{b} + \frac{\varepsilon_0}{\varepsilon}}$$

$$\phi_2 = -\left(\frac{\rho_{fc} x^2}{2\varepsilon}\right) + \frac{\left(\frac{\rho_{fc} bx}{2\varepsilon}\right)\left(\frac{\varepsilon_0}{\varepsilon} + \frac{2a}{b}\right)}{\frac{a}{b} + \frac{\varepsilon_0}{\varepsilon}} \tag{16.27}$$

and the corresponding **E** fields are

$$\mathbf{E}_1 = \mathbf{i}_x \left[\frac{\frac{\rho_{fc} b}{2\varepsilon}}{\frac{a}{b} + \frac{\varepsilon_0}{\varepsilon}}\right]$$

and

$$\mathbf{E}_2 = \mathbf{i}_x \left[\frac{\left(\frac{\rho_{fc} x}{\varepsilon} - \frac{\rho_{fc} b}{2\varepsilon}\right)\left(\frac{\varepsilon_0}{\varepsilon} + \frac{2a}{b}\right)}{\frac{a}{b} + \frac{\varepsilon_0}{\varepsilon}}\right] \tag{16.28}$$

Solution in charge decay. The enclosed control volume in the system is shown in Figure 16.1, and the surface area normal to the x-direction is taken as S. It should be noted that J_1 in the region (1)-free space is zero, because the conductivity is zero. Hence from Eq. (16.3), by integration with respect to x, we get

$$J_2 S = -\frac{d}{dt}\int_0^b \rho_{fc} S\, dx = -\frac{d}{dt}(\rho_{fc} Sb)$$

Also

$$J_2 = -\sigma E_2 \text{ at } x = 0$$

$$\therefore \quad \sigma S E_2 = \frac{d}{dt}(\rho_{fc} Sb) \tag{16.29}$$

Combining Eq. (16.29) with (16.28) for $x = 0$, we get

$$\frac{\frac{d\rho_{fc}}{dt} + \left(\frac{\sigma\rho_{fc}}{2\varepsilon}\right)\left(\frac{\varepsilon_0}{\varepsilon} + \frac{2a}{b}\right)}{\frac{a}{b} + \frac{\varepsilon_0}{\varepsilon}} = 0 \tag{16.30}$$

whose solution is obtained directly as

$$\rho_{fc} = \rho_0 \exp\left(-\frac{t}{\tau_1}\right) \tag{16.31}$$

where ρ_0 is the initial charge density, and τ_1—the new relaxation time which is given by

$$\tau_1 = \left(\frac{\varepsilon}{\sigma}\right)\left[\frac{2\left(\frac{\varepsilon_0}{\varepsilon} + \frac{a}{b}\right)}{\frac{\varepsilon_0}{\varepsilon} + \frac{2a}{b}}\right] \tag{16.32}$$

Note that $(\varepsilon/\sigma) = \tau$, which is our conventional relaxation time. Writing $a + b = d$ and $(b/d) = h$, then

$$\frac{\tau_1}{\tau} = \frac{2\,[h + \varepsilon_r\,(1 - h)]}{h + 2\varepsilon_r\,(1 - h)} \tag{16.33}$$

So we see that the maximum value of this ratio (τ_1/τ) is 2 and is independent of either ε_r or σ.

We can similarly solve the composite Laplacian and Poissonian field problem for the potential distribution in two or even three dimensions.

16.2.2 Media with Non-uniform Properties

In such a medium (at rest) in which the conductivity and permittivity are functions of space, a steady-state volume charge density can exist. The interface between two dissimilar media is a 'special case' of this type and the interface gradients of ε and σ are 'singularities' and hence the surface charges accumulate on the surfaces of the uniformly conducting media.

When σ and ε are functions of position in a stationary medium, the potential equation (16.9) becomes

$$\nabla \cdot (\sigma \nabla \phi) + \frac{\partial}{\partial t}[\nabla \cdot (\varepsilon \nabla \phi)] = 0 \tag{16.34}$$

which cannot take the simplified form of Eq. (16.11), because σ and ε are no longer constants and they vary with the coordinate directions. We again consider a simple one-dimensional example of this type.

16.2.2.1 Example 1: A model with non-uniform properties

A material of non-uniform properties (i.e. σ and ε are functions of space coordinates) is bound by two plane parallel electrodes [Figure 16.2(a)]. An external current source drives a current through this material in the x-direction. For simplicity, we also assume σ and ε to be functions of x only. Physically such a system is created when the two electrodes with temperature difference are placed in an organic liquid. In the present one-dimensional problem, Eq. (16.34) becomes

$$\frac{\partial}{\partial x}\left[\sigma E_x + \frac{\partial}{\partial t}(\varepsilon E_x)\right] = 0 \tag{16.35}$$

where $(\text{grad}\ \phi)_x = -E_x$. Integrating the above equation with respect to x

$$\left(\frac{\partial}{\partial t}\right)(\varepsilon E_x) + \sigma E_x = f(t) \tag{16.36}$$

The physical interpretation of the above equation is that the sum of the displacement current and the conduction current is same all over the y–z planes. The function $f(t)$ is the current density $[i(t)/A]$ where A is the surface area of the electrodes. Hence Eq. (16.36) becomes

$$\left(\frac{\partial}{\partial t}\right)(\varepsilon E_x) + \sigma E_x = \frac{i(t)}{A} \tag{16.37}$$

Let us consider a sinusoidal time-varying driving current

$$i(t) = \text{Re}\left[\hat{I}\ \exp(j\omega t)\right]$$

$$\therefore \quad E_x(t) = \text{Re}\left[\hat{E}_x(x)\ \exp(j\omega t)\right] \tag{16.38}$$

$$\therefore \quad \hat{E}_x = \frac{\hat{I}}{A(j\omega\varepsilon + \sigma)} \tag{16.39}$$

Next we find the charge density from Eq. (16.2), which in this case becomes

$$\rho_{\text{fc}} = \nabla \cdot \mathbf{D} = \nabla \cdot (\varepsilon \mathbf{E}) = \left(\frac{\partial}{\partial x}\right)(\varepsilon E_x)$$

$$= \left(\frac{\partial}{\partial x}\right)\left[\frac{\varepsilon \hat{I}\ \exp(j\omega t)}{A(j\omega\varepsilon + \sigma)}\right]$$

$$\therefore \quad \hat{\rho}_{\text{fc}} = -\left(\frac{\varepsilon \hat{I}}{A}\right)\left[\frac{j\omega\left(\frac{d\varepsilon}{dx}\right) + \left(\frac{d\sigma}{dx}\right)}{(j\omega\varepsilon + \sigma)^2}\right] + \frac{\hat{I}\left(\frac{d\varepsilon}{dA}\right)}{A(j\omega\varepsilon + \sigma)} \tag{16.40}$$

For this problem, let us assume that the conductivity varies linearly with respect to x (Figure 16.2b), and ε_{r} is constant, then

$$\sigma(x) = \sigma_0 + \left(\frac{\sigma_1}{l}\right)x \tag{16.41}$$

Hence

$$\left|\hat{E}_x\right| = \frac{\left|\hat{I}\right|}{A(\omega^2\varepsilon^2 + \sigma^2)^{1/2}}$$

and

$$\left|\hat{\rho}_{\text{fc}}\right| = \frac{\varepsilon\left|\hat{I}\right|\sigma_1}{Al(\omega^2\varepsilon^2 + \sigma^2)^{1/2}} \tag{16.42}$$

E_x and ρ_{fc} distributions are shown in the Figures 16.2(c) and (d) respectively. E_x is roughly inversely proportional to x, and the charge density is inversely proportional to x^2. It should be noted that the free charge accumulation is maximum, when $\omega = 0$, i.e. the current is constant. The charge accumulation due to a non-uniform conductivity occurs in oil-immersed HVDC cables used for transmission of power.

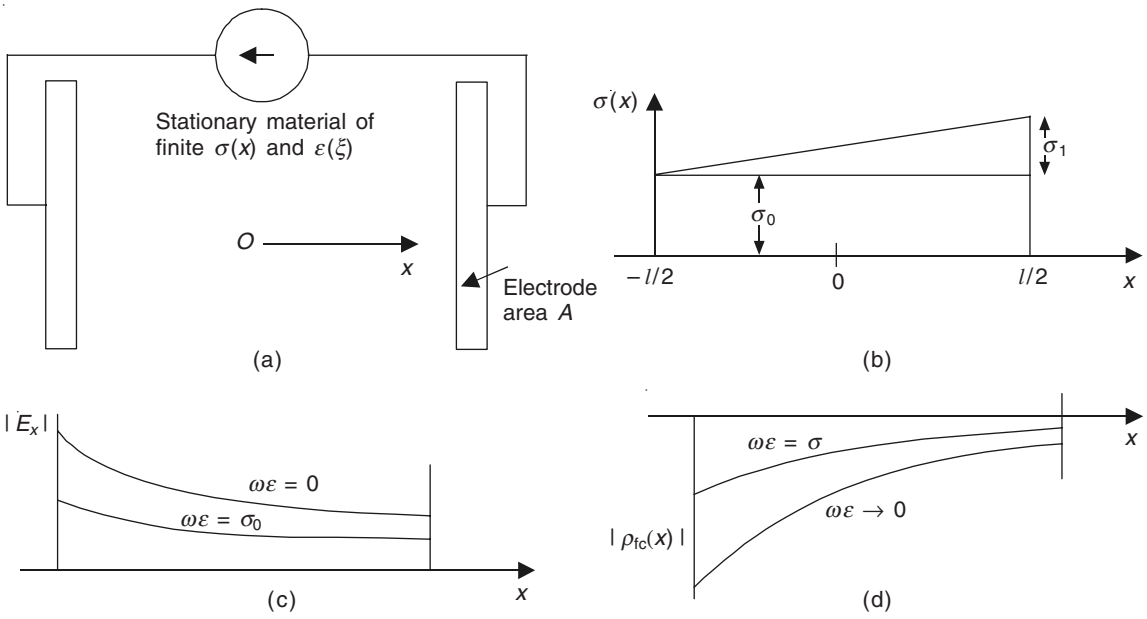


Figure 16.2 (a) A partially conducting material, bounded by plane electrodes; (b) σ distribution between the electrodes; (c) E_x distribution; and (d) ρ_{fc} distribution.

16.3 CHARGE RELAXATION WHEN THERE IS STEADY MOTION

The next aspect of charge relaxation, which we shall look at, is the effect of material motion on relaxation. Now the material properties of the media (i.e. σ and ε) are taken as constant in the bulk of the material, and change only at the interface surfaces. We again start with the general operational equation (16.9), which is

$$\nabla \cdot (\sigma \nabla \phi) + \nabla \cdot [\mathbf{v} \nabla \cdot (\varepsilon \nabla \phi)] = - \left(\frac{\partial}{\partial t} \right) [\nabla \cdot (\varepsilon \nabla \phi)] \tag{16.9}$$

To express the above equation in terms of the free-charge density, we use the equations

$$\mathbf{E} = - \nabla \phi \qquad \text{and} \qquad \nabla \cdot \mathbf{D} = \nabla \cdot (\varepsilon \mathbf{E}) = \rho_{fc}, \text{ and obtain}$$

$$\left(\frac{\sigma}{\varepsilon} \right) \rho_{fc} + \nabla \cdot (\rho_{fc} \mathbf{v}) + \left(\frac{\partial \rho_{fc}}{\partial t} \right) = 0 \tag{16.43}$$

The important case of this type is that of incompressible medium in which the material density remains constant, and the net flux of the material in a given region is zero, which means that, for the incompressible material, the divergence of the velocity must be zero (analogy to div $\mathbf{B} = 0$). Hence

$$\text{div } \mathbf{v} = 0 \tag{16.44}$$

Hence expanding the second term of Eq. (16.43)

$$\begin{aligned} \nabla \cdot (\rho_{fc} \mathbf{v}) &= \mathbf{v} \cdot (\nabla \rho_{fc}) + \rho_{fc} (\nabla \cdot \mathbf{v}) \\ &= \mathbf{v} \cdot (\nabla \rho_{fc}), \text{ for incompressible medium} \end{aligned} \tag{16.45}$$

Equation (16.43) can be rewritten as (for this case)

$$\left(\frac{\partial \rho_{fc}}{\partial t}\right) + \left(\frac{\sigma}{\varepsilon}\right)\rho_{fc} + \mathbf{v} \cdot (\nabla \rho_{fc}) = 0 \tag{16.46}$$

This equation gives the same relaxation condition as Eq. (16.11), except that the charge relaxation now occurs with respect to the frame of the moving medium, i.e. the material motion transports the free charge as it relaxes with the time-constant (ε/σ). We shall now describe a device in which the charge is transported from a region of one potential to a region of another. The time required for a given initial distribution of charge to relax (with respect to the medium) is (ε/σ), and the time necessary for transporting the charge over a distance l with velocity $\mathbf{v}$ is (l/v). Thus the electric Reynold's number R_e:

$$R_e = \frac{\varepsilon/\sigma}{l/\mathrm{v}} = \frac{\varepsilon \mathrm{v}}{\sigma l} >> 1 \tag{16.47}$$

must be much greater than unity, if the convection is to compete with the relaxation process in deciding the location of the volume charge density.

16.3.1 Van-de-Graff Generator

This is used for generating high voltage dc power by transporting free charge against an electric field. A highly simplified model is shown in Figure 16.3. Essentially it consists of a continuous belt made of slightly conducting material, and driven by rollers at constant velocity which is

$$\mathbf{v} = \mathbf{i}_z \mathrm{v}$$

The electrode at $z = 0$ feeds positive ions on to the surface of the belt, and the electrode at

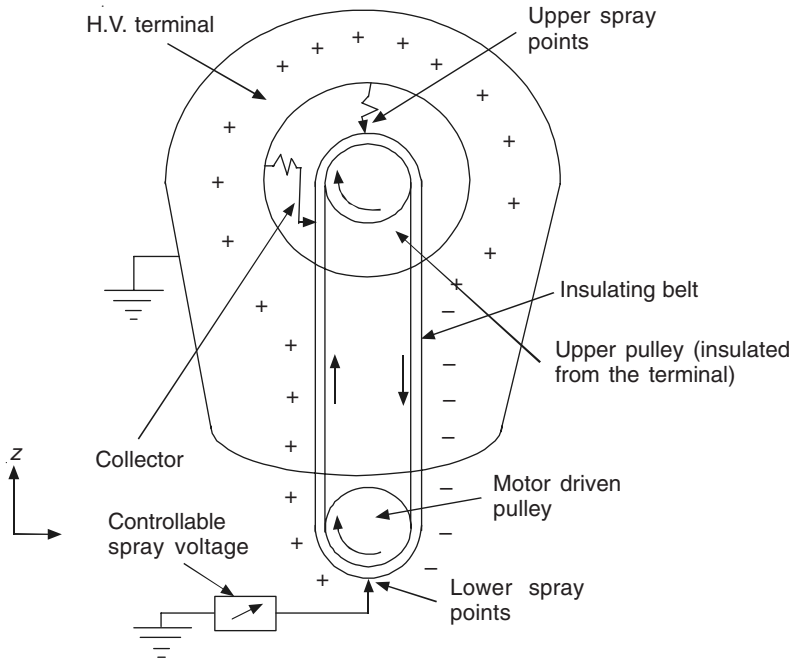


Figure 16.3 Section of a Van-de-Graff generator.

$z = l$ removes the positive charges. For our analysis, we consider these two electrodes under open-circuited condition. The belt material has constant σ and ε. The positive ions put on the belt are immobile.

The simplifying assumptions for the analysis are:

1. We consider only the portion of the belt that carries charge from
$$z = 0 \text{ to } z = l$$
2. **E** and **J** have z-component only, i.e.
$$\mathbf{E} = \mathbf{i}_z E_z \quad \text{and} \quad \mathbf{J} = \mathbf{i}_z J_z \tag{16.48}$$
3. All variables are functions of z alone.
4. For this one-dimensional model, the effect of the applied positive ions and the induced charges in the belt can be represented by an effective free-charge density $\rho_{fc}(z)$ which also is a function of z only.
5. The boundary condition at $z = 0$ is: $\rho_{fc} = \rho_0$ (16.49)
6. The system is operating in steady state.

With σ constant, the constituent relation for the transport of free charges is Eq. (16.5), i.e.

$$\mathbf{J} = \sigma\mathbf{E} + \rho_{fc}\mathbf{v} \tag{16.5}$$

Gauss' theorem [Eq. (16.2)] gives

$$\varepsilon\left(\frac{dE_z}{dz}\right) = \rho_{fc} \tag{16.50}$$

For this problem, the open-circuit operation implies that

$$\mathbf{J} = 0 \tag{16.51}$$

Differentiating Eq. (16.5) with respect to z, and substituting from Eq. (16.50) to eliminate E_z, we get

$$\mathrm{v}\left(\frac{d\rho_{fc}}{dz}\right) + \left(\frac{\sigma}{\varepsilon}\right)\rho_{fc} = 0 \tag{16.52}$$

The general solution of this equation is

$$\rho_{fc} = C_1 \exp\left[-(\sigma/\varepsilon)(z/\mathrm{v})\right] \tag{16.53}$$

Using the boundary condition at $z = 0$ to evaluate C_1, we get

$$\rho_{fc} = \rho_0 \exp\left(\frac{-z}{R_e l}\right)$$

where

$$R_e = \frac{\varepsilon \mathrm{v}}{\sigma l} \tag{16.54}$$

$\therefore$ From Eq. (16.5),

$$E_z = -\left(\frac{\mathrm{v}}{\sigma}\right)\rho_{fc} = -\left(\frac{\mathrm{v}}{\sigma}\right)\rho_0 \exp\left(\frac{-z}{R_e l}\right) \tag{16.55}$$

and hence the open-circuit (o.c.) voltage is

$$V_{\text{o.c.}} = -\int_{z=0}^{z=1} E_z\,dz = \left(\frac{R_e^2 l^2}{\varepsilon}\right)\rho_0\left[1-\exp\left(-\frac{1}{R_e}\right)\right] \tag{16.56}$$

16.4 CHARGE RELAXATION WITH MOTION AND SINUSOIDAL EXCITATION

We now consider a problem of charge relaxation, when it is associated with motion and sinusoidal excitation. This example shows the distinguishing features of relaxation and magnetic diffusion, and also an application of the phenomenon to measure the velocity of a moving medium.

A slightly conducting thin slab (of characteristic properties σ, ε) is moving to the right with a constant velocity **v** (as shown in Figure 16.4) between two plane parallel electrodes,

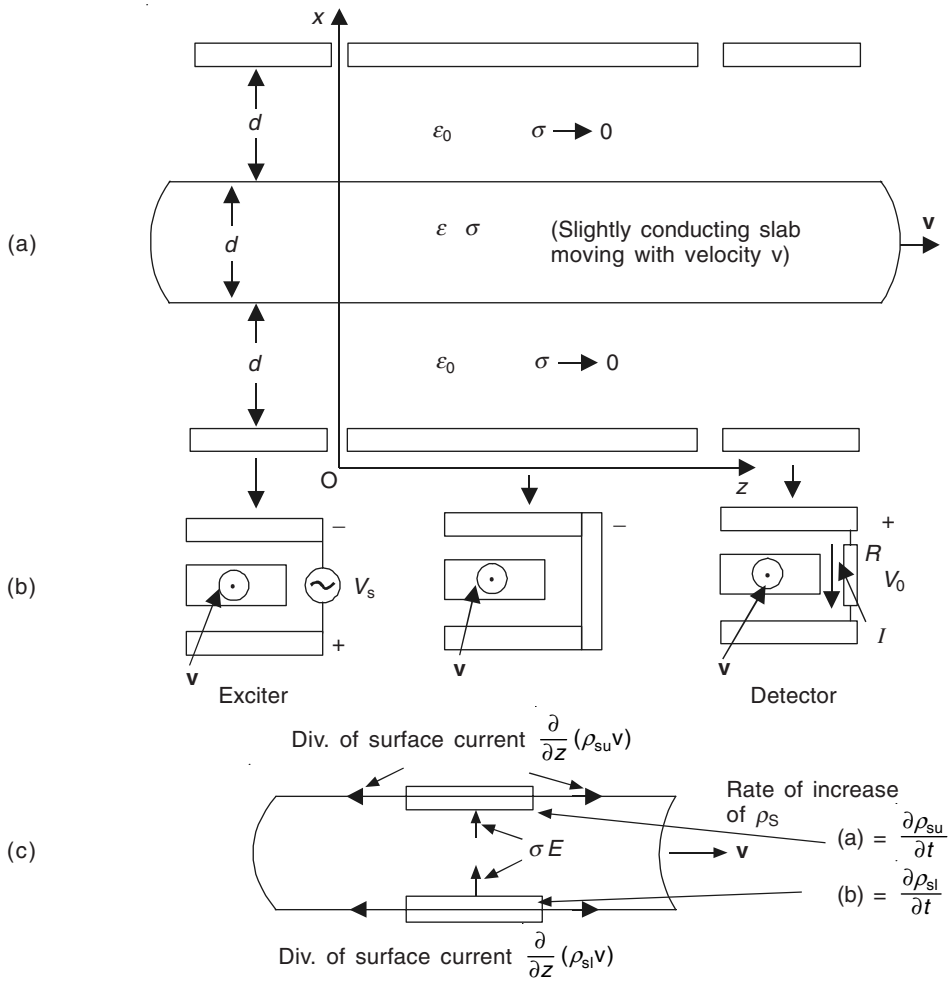


Figure 16.4 (a) A slightly conducting thin slab, moving to the right with constant velocity **v** through electric fields imposed and constrained as shown in (b); (c) conduction currents normal to the surfaces produce a surface charge accumulation as shown.

which are sub-divided into three sections with different terminations as shown in Figure 16.4(b). In the section to the left, the electrodes are excited by a sinusoidally varying potential.

Since the slab is insulated from the electrodes by the air-gaps between them, the charge would relax to the surface so as to prevent the externally imposed field from entering it. Hence as the slab leaves the exciter region of the electrodes at $z = 0$, there would be a sinusoidally varying surface-charge density ($\pm \rho_0$) on the upper and the lower surfaces of the slab respectively. As the slab passes through the region $0 < z < l$, the surface charges, which had been induced in the exciter region, tend to relax, and by the time they reach the detector region [the right section of the electrode in Figure 16.4(b)], the surface-charge densities get considerably attenuated. The charges, which remain on the surface by this time, induce image charges on the electrodes, which result a current through the resistance R, thereby generating an output signal V_0. Because of the charge relaxation, this output signal can be used as a measure of the velocity of the block.

There is no charge inside the moving slab, except on its surfaces as the normal conduction current leads to an accumulation of charges on the surface, which is convected by the motion of the slab. Hence, the conduction current on the upper surface of the block is

$$= \sigma E = \left(\frac{\partial \rho_{su}}{\partial t} \right) + v \left(\frac{\partial \rho_{su}}{\partial z} \right) \tag{16.57}$$

where the surface-charge density ρ_{su} is given by the equation

$$\rho_{su} = \varepsilon_0 E_u - \varepsilon E \tag{16.58}$$

Similarly for the lower surface of the block

$$- \sigma E = \left(\frac{\partial \rho_{sl}}{\partial t} \right) + v \left(\frac{\partial \rho_{sl}}{\partial z} \right) \tag{16.59}$$

where

$$\rho_{sl} = \varepsilon E - \varepsilon_0 E_l \tag{16.60}$$

In general, the electric field in the slab and in the air-gaps above and below it, must satisfy the electric field equations with $\rho_{fc} = 0$. To simplify the problem, we assume a simple one-dimensional model, assuming $\mathbf{E} = \mathbf{i}_x E(z, t)$, and

$$d(E_u + E + E_l) = 0 \tag{16.61}$$

because the integral of the field along a line joining the plates must be zero.

This one-dimensional model ignores the effect of the z-component of $\mathbf{E}$ which does exist and can be evaluated from the condition $\nabla \times \mathbf{E} = 0$. At $z = 0$, we assume that the exciter has induced the surface charges

$$\rho_{sl} = - \rho_{su} = \rho_0 \sin \omega t \tag{16.62}$$

$\therefore$ From Eq. (16.61),

$$E = - (E_u + E_l) = - 2E_u \tag{16.63}$$

(from symmetry considerations $E_u = E_l$)

and from Eq. (16.58),

$$E = - \frac{2\rho_{su}}{\varepsilon_0 + 2\varepsilon} \tag{16.64}$$

On substituting for E in Eq. (16.57),

$$\left(\frac{\partial \rho_{su}}{\partial t}\right) + v\left(\frac{\partial \rho_{su}}{\partial z}\right) + \left(\frac{2\sigma}{\varepsilon_0 + 2\varepsilon}\right)\rho_{su} = 0 \tag{16.65}$$

Similarly the corresponding equation for the lower surface can be derived ($\rho_{su} = -\rho_{sl}$ and $E_u = E_l$).

We assume the steady-state solution of the form

$$\rho_{su} = \mathrm{Re}\left[\hat{\rho}_{su}(z) \exp(j\omega t)\right] \tag{16.66}$$

and substituting in Eq. (16.65), gives

$$\left(\frac{\partial \hat{\rho}_{su}}{\partial z}\right) + \left[\frac{2\sigma}{(\varepsilon_0 + 2\varepsilon)v} + \frac{j\omega}{v}\right]\rho_{su} = 0 \tag{16.67}$$

The solution to this equation is

$$\rho_{su}(z, t) = \rho_0 \exp\left(-\frac{z}{R_e l}\right)\sin \omega\left(t - \frac{z}{v}\right)$$

where

$$R_e = \frac{(\varepsilon_0 + 2\varepsilon)v}{2\sigma l} \tag{16.68}$$

where R_e is the electric Reynold's number, and gives a measure of the exponential damping effect, similar to that shown in Figure 15.2 of Section 15.2. Thus the relaxing charge appears as a damped wave propagating to the right with a phase velocity v and an attenuating factor whose exponent is $(-1/R_e l)$. It should be noted that the relaxation wave propagates by virtue of the material movement (as distinct from the case of magnetic diffusion where the propagating wave existed even in stationary medium), and the attenuation was determined by the combination of the material properties and the velocity of the material.

16.5 TRAVELLING WAVE CHARGE RELAXATION IN A MOVING CONDUCTOR

The physical configuration of the problem is shown in Figure 16.5. This is a slightly conducting, semi-infinite block moving to the right with a velocity v, just below a segmented electrode, which supports a travelling wave of potential given by

$$V = \mathrm{Re}\left[\hat{V} \exp\{j(\omega t - kz)\}\right] \tag{16.69}$$

The conducting material [region (2)] has the material characteristics ε and σ which are constant and uniform, and moves to the right with constant velocity $\mathbf{v} = \mathbf{i}_z v$. For this material, there is no bulk free-charge inside. The surface charges exist on the interface plane $x = 0$ and its distribution is determined by the motion of the medium and the velocity of the travelling wave of potential.

Since there is no free charge either in region (1) or in region (2), the operational equation for the potential distribution for both the regions will be Laplacian, i.e.

$$\nabla^2 \phi = \frac{\partial^2 \phi}{dx^2} + \frac{\partial^2 \phi}{\partial y^2} = 0 \tag{16.70}$$

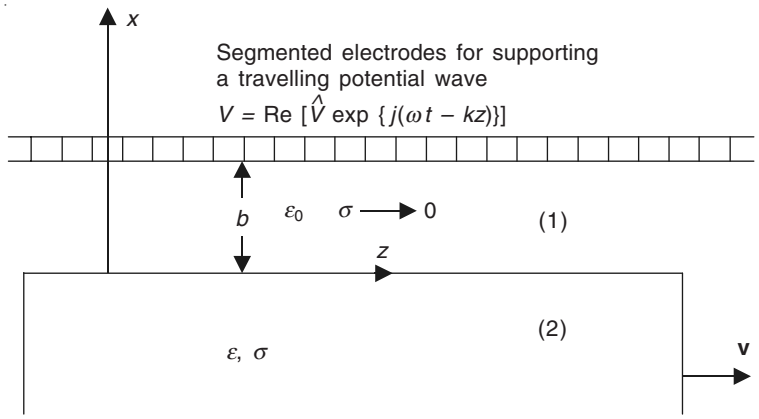


Figure 16.5 A semi-infinite block of slightly conducting material moving to the right with a constant velocity $\mathbf{v} = \mathbf{i}_z v$.

As can be seen from Figure 16.5, the problem would be two-dimensional, and hence there would be no variation in the y-direction (i.e. $\partial/\partial y \equiv 0$). To match the condition of the source potential as given by Eq. (16.69), we assume the potential to be of the form

$$\phi = \text{Re}\left[\hat{\phi} \exp\{j(\omega t - kz)\}\right] \tag{16.71}$$

in both the regions (1) and (2).

In the region (1), the potential function will take the form

$$\hat{\phi}_1 = A \sinh(kx) + B \cosh(kx) \tag{16.72}$$

where A and B are arbitrary constants to be determined by the boundary conditions.

In the region (2), the potential function will take the form

$$\hat{\phi}_2 = C \exp(+kx) \tag{16.73}$$

This assumes that k is +ve, since the potential will not go to infinity as $x \to (-\infty)$.

We need three boundary conditions to evaluate the three unknowns A, B, and C.

Boundary conditions

1. At $x = b$, $V = \text{Re}\left[\hat{V} \exp\{j(\omega t - kz)\}\right]$ (16.69)
2. At $x = 0$, $\hat{\phi}_1(0) = \hat{\phi}_2(0)$ (16.74)
3. At $x = 0$, $\rho_{S0} = \varepsilon_0 E_{1x} - \varepsilon E_{2x} = -\varepsilon_0\left(\frac{\partial \phi_1}{\partial x}\right) + \varepsilon\left(\frac{\partial \phi_2}{\partial x}\right)$ (16.75)

(conservation of the charge on the interface surface).

This condition, expressed in terms of the conduction current normal to the interface gives an increase in the surface charge (measured in a frame moving with the material) as

$$-\sigma\left(\frac{\partial \phi_2}{\partial x}\right) = \left[\left(\frac{\partial}{\partial t}\right) + \text{v}\left(\frac{\partial}{\partial z}\right)\right]\left[-\varepsilon_0\left(\frac{\partial \phi_1}{\partial x}\right) + \varepsilon\left(\frac{\partial \phi_2}{\partial x}\right)\right] \tag{16.76}$$

For the assumed travelling wave solution of Eq. (16.71), this becomes

$$-\sigma\left(\frac{\partial\hat{\phi}_2}{\partial x}\right)=j(\omega-k\mathrm{v})\left[-\varepsilon_0\left(\frac{\partial\hat{\phi}_1}{\partial x}\right)+\varepsilon\left(\frac{\partial\hat{\phi}_2}{\partial x}\right)\right] \tag{16.77}$$

$\therefore$ Using the boundary condition equations of (16.69), (16.74), and (16.77), we get

$$\begin{aligned} A\sinh(kb)+B\cosh(kb) &= \hat{V} \\ B-C &= 0 \\ AjS-C\left[1+jS\left(\frac{\varepsilon}{\varepsilon_0}\right)\right] &= 0 \end{aligned} \tag{16.78}$$

where S = the normalized frequency measured in the moving frame of the material, normalized to the relaxation time

$$=(\omega-k\mathrm{v})\frac{\varepsilon_0}{\sigma} \tag{16.79}$$

Evaluating the unknowns, we obtain

$$\begin{aligned} \phi_1 &= \mathrm{Re}\left(\frac{V}{D}\right)\left[\left(1+jS\frac{\varepsilon}{\varepsilon_0}\right)\sinh(kx)+jS\cosh(kx)\right]\left[\exp\{j(\omega t-kz)\}\right] \\ \phi_2 &= \mathrm{Re}\left(\frac{\hat{V}}{D}\right)\left[jS\exp(kz)\exp\{j(\omega t-kz)\}\right] \end{aligned} \tag{16.80}$$

where

$$D=\left[\left\{1+jS\left(\frac{\varepsilon}{\varepsilon_0}\right)\right\}\sinh(kb)+jS\cosh(kb)\right] \tag{16.81}$$

For the range of values of S, there are two limiting cases:

1. $S = 0$, i.e.

$$\begin{aligned} \phi_1 &= \mathrm{Re}\,\hat{V}\left[\frac{\sinh(kx)}{\sinh(kb)}\right]\left[\exp\left\{j(\omega t-kz)\right\}\right] \\ \phi_2 &= 0 \end{aligned} \tag{16.82}$$

i.e. the medium behaves as perfectly conducting, and there is no penetration of the **E**-field in the moving medium.

2. When $S \to \infty$ (i.e. very large),

$$\phi_1 = \mathrm{Re}\,\hat{V}\left[\frac{\sinh(kx)+\left(\frac{\varepsilon}{\varepsilon_0}\right)\cosh(kx)}{\sinh(kb)+\left(\frac{\varepsilon}{\varepsilon_0}\right)\cosh(kb)}\right]\times\left[\exp\left\{j(\omega t-kz)\right\}\right]$$

$$\phi_2 = \mathrm{Re}\,\hat{V}\left[\frac{\exp(kx)\exp\left\{j(\omega t-kz)\right\}}{\sinh(kb)+\left(\frac{\varepsilon}{\varepsilon_0}\right)\cosh(kb)}\right] \tag{16.83}$$

In this case, the electric field completely penetrates the moving medium, i.e. the moving block behaves like a perfectly insulating material with no free charges at the surface, as can be seen by putting $x = 0$ in the above equations. The reason is that in this high frequency limit, one time-period of the excitation is not sufficient time for appreciable free-charge to relax to the surface. Since the magnitude of the parameter S depends directly on both the frequency ω of the imposed travelling wave excitation as well as on the velocity v of the medium, the insulating behaviour of the medium (i.e. the field penetration in the block) can be obtained by controlling either of these parameters, i.e. either increasing the frequency ω of the imposed excitation or increasing the velocity v of the moving block.

16.6 SOME COMMENTS ON THE CHARGE DECAY IN CONDUCTING MEDIA

Recently, a number of investigators have raised questions about the validity of the relaxation equation [Eqs. (16.11) and (16.12)] for the conductors. It has been argued that for good conductors, this equation gives too small a value of the relaxation time τ, which can be practically correct. It has been argued that the cause for this discrepancy is the assumption of constant conductivity ($=\sigma$) under the conditions of varying electric field. A more correct approach would be to take the 'mobility' (i.e. the magnitude of the drift velocity v per unit electric field) as constant instead, i.e.

$$\mu\,(\text{mobility}) = \frac{|\text{v}|}{E}$$

(In this section μ stands for the mobility of the charges, and should not be confused with the notation for the magnetic permeability used elsewhere in this book). The electrical conductivity is the sum of the electron and hole contributions

$$\sigma = ne\mu_e + pe\mu_e$$

where n and p are the concentrations of the electrons and holes in the medium (and e the electron charge).

We have merely stated this point here (in a textbook of this level) as a matter of interest and do not intend to pursue the subject in depth, which is really the domain of specialists.

PROBLEMS

16.1 Calculate the relaxation time for ethyl alcohol, for which $\varepsilon_r = 26$ and $\sigma = 3 \times 10^{-4}$ mho/m.

Ans.: $\tau \approx 10^{-6}$ sec.

16.2 In the problem of Section 16.2.2.1 (media with non-uniform properties), if the permittivity ε varied linearly with x, along with the conductivity, then

$$\varepsilon(x) = \varepsilon_1 + \frac{\varepsilon_2}{l}\,x, \qquad \sigma(x) = \sigma_1 + \frac{\sigma_2}{l}\,x \qquad \text{[referring to Figure 16.2(a)]}$$

Show that the free charge density is given by

$$\hat{\rho}_{fc} = -\frac{I}{a}\left[\frac{\left(\varepsilon_1 + \frac{\varepsilon_2}{l}x\right)\left(j\omega\frac{\varepsilon_2}{l} + \frac{\sigma_2}{l}\right)}{\left(j\omega\varepsilon + \sigma\right)^2} + \frac{\frac{\varepsilon_2}{l}}{\left(j\omega\varepsilon + \sigma\right)}\right]$$

Write down the expression for the free charge density, when the permittivity $\varepsilon(x)$ changes linearly with x, while keeping the conductivity σ constant.

17 Electromagnetic Waves

17.1 INTRODUCTION

One of the most brilliant achievements of Maxwell's theory of electromagnetism was the prediction of the existence of electromagnetic waves. This was confirmed experimentally twenty years later by the work of the German physicist 'Heinrich Hertz (1857–Jan 1894) who made this very significant discovery in the short-span of his life of 37 years. Hertz was a very bright student who started studying engineering but changed over to physics at the University of Berlin under Hermann von Helmholtz and Gustav R. Kirchhoff, who were of the mechanistic school of thought (i.e. action at a distance) on the continent as against the field concept propounded by Maxwell in Britain. Up to mid-1880s, Hertz with his mechanistic background, found the field theory and the vector approach from the Scottish/English schools rather baffling. Yet Hertz got convinced of the reality of the electric and the magnetic wave radiation. His first experiment, with a circular loop of wire with a tiny gap between the ends detecting a spark from a discharge in a nearby induction coil, with no physical contact between the two, was the first properly understood transmission and reception of electromagnetic radiation through air.

In the present chapter, we shall start our study with the simplest type of wave, i.e. the uniform plane wave, and then develop the subject matter further. Initially we shall deal with the presence of these waves in free space, and then consider other types of media, such as, conductors, dielectrics, etc. At this stage of our study, we shall not deal with the sources of the electromagnetic waves, till Chapter 19.

Our starting point would be, as in the case of the eddy currents and the magnetic diffusion, the Maxwell's equations, the only difference being that now we ignore the conduction current term of Eq. (12.11) or (12.18) from Chapter 12 (we shall make this simplifying assumption for all loss-less ideal media). Thus our operating equation would be the first limiting case of the vector Helmholtz equation, i.e. Eqs. (12.38) which are

$$\nabla^2\mathbf{H} + \beta^2\mathbf{H} = 0 \tag{12.38a}$$

$$\nabla^2\mathbf{E} + \beta^2\mathbf{E} = 0 \tag{12.38b}$$

where $\beta^2 = \omega^2 \mu_0 \varepsilon_0 = \omega^2/c^2$, $c = 1/\sqrt{\mu_0 \varepsilon_0}$, in free space. Sometimes in place of β we use the notation k, which is also called the wave number of the wave.

17.2 UNIFORM PLANE ELECTROMAGNETIC WAVE IN FREE SPACE

In fact, we have already met with and discussed the nature and the behaviour of these waves earlier in Chapter 12, Section 12.6. For that analysis, we had simplified the associated mathematics very significantly by rotating the coordinate system such that the **E** vector coincided with the x-axis. This adjustment maintained the complete generality of the problem and simplified the analysis considerably. However such simplification is not always possible. There are configurations and considerations when it is not possible to fiddle with the coordinate system to simplify the problem. So now, we keep our analysis general and make no simplifying arrangements at the start of our problem (as we had done in Section 12.6). As before, we start with the Maxwell's equations, neglecting the conduction current **J**; i.e.

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad \nabla \cdot \mathbf{B} = 0 \tag{17.1}$$

$$\nabla \times \mathbf{H} = -\frac{\partial \mathbf{D}}{\partial t}, \quad \nabla \cdot \mathbf{D} = 0 \qquad \text{(there is no charge in the medium)} \tag{17.2}$$

and the constituent relations, relevant to our present problem

$$\mathbf{B} = \mu_0 \mu_r \mathbf{H} = \mu \mathbf{H} \qquad \text{and} \qquad \mathbf{D} = \varepsilon_0 \varepsilon_r \mathbf{E} = \varepsilon \mathbf{E} \tag{17.3}$$

As we have seen in Section 12.3, these equations reduce to

$$\nabla^2 \mathbf{H} = \mu \varepsilon \left(\frac{\partial^2 \mathbf{H}}{\partial t^2} \right) \tag{17.4a}$$

$$\nabla^2 \mathbf{E} = \mu \varepsilon \left(\frac{\partial^2 \mathbf{E}}{\partial t^2} \right) \tag{17.4b}$$

which are wave equations.

We now define the uniform plane electromagnetic waves (in words):

*These are waves in which the field vectors **E** and **H** have constant values of magnitude and phase on planes perpendicular to the direction of propagation.*

Let us assume the direction of propagation to be the $+z$-axis of our coordinate system. Then, by definition, **E** and **H** will have constant values on all planes perpendicular to the z-axis.

$$\therefore \qquad \frac{\partial \mathbf{E}}{\partial x} = \frac{\partial \mathbf{E}}{\partial y} = \frac{\partial \mathbf{H}}{\partial x} = \frac{\partial \mathbf{H}}{\partial y} = 0 \tag{17.5a}$$

Note that, so far, we have imposed no constraints on the vectors **E** and **H**. However it will come out later that these vectors **E** and **H** can have no components in the z-direction, i.e.

$$E_z = 0, \quad H_z = 0 \tag{17.5b}$$

(At this stage we impose no such condition.)

∴ The wave equations derived from the Maxwell's equations reduce to

$$\nabla^2\mathbf{H} \equiv \frac{\partial^2\mathbf{H}}{\partial z^2} = \mu\varepsilon\left(\frac{\partial^2\mathbf{H}}{\partial t^2}\right) \tag{17.6}$$

which implies

$$\frac{\partial^2 H_x}{\partial z^2} = \mu\varepsilon\left(\frac{\partial^2 H_x}{\partial t^2}\right), \quad \frac{\partial^2 H_y}{\partial z^2} = \mu\varepsilon\left(\frac{\partial^2 H_y}{\partial t^2}\right), \quad \text{and} \quad \frac{\partial^2 H_z}{\partial z^2} = \mu\varepsilon\left(\frac{\partial^2 H_z}{\partial t^2}\right) \tag{17.7}$$

and

$$\nabla^2\mathbf{E} \equiv \frac{\partial^2\mathbf{E}}{\partial z^2} = \mu\varepsilon\left(\frac{\partial^2\mathbf{E}}{\partial t^2}\right) \tag{17.8}$$

which implies

$$\frac{\partial^2 E_x}{\partial z^2} = \mu\varepsilon\left(\frac{\partial^2 E_x}{\partial t^2}\right), \quad \frac{\partial^2 E_y}{\partial z^2} = \mu\varepsilon\left(\frac{\partial^2 E_y}{\partial t^2}\right), \quad \text{and} \quad \frac{\partial^2 E_z}{\partial z^2} = \mu\varepsilon\left(\frac{\partial^2 E_z}{\partial t^2}\right) \tag{17.9}$$

i.e. six similar equations which will reduce to four similar equations.

To start with, we solve for $H_{x:}$. The expression for the solution is of the form

$$H_{x:} = f_1(z - \mathrm{u}t) + f_2(z + \mathrm{u}t) \tag{17.10}$$

where u is independent of t.

Proof. Let

$$(z - \mathrm{u}t) = m, \quad (z + \mathrm{u}t) = n \tag{17.11}$$

then

$$\frac{\partial m}{\partial z} = 1, \ \frac{\partial n}{\partial z} = 1, \ \frac{\partial m}{\partial t} = -\mathrm{u}, \ \frac{\partial n}{\partial t} = +\mathrm{u} \tag{17.12}$$

∴

$$\frac{\partial H_x}{\partial z} = \left(\frac{\partial f_1}{\partial m}\right)\left(\frac{\partial m}{\partial z}\right) + \left(\frac{\partial f_2}{\partial n}\right)\left(\frac{\partial n}{\partial z}\right) = \left(\frac{\partial f_1}{\partial m}\right) + \left(\frac{\partial f_2}{\partial n}\right) \tag{17.13}$$

and

$$\frac{\partial^2 H_x}{\partial z^2} = \left(\frac{\partial}{\partial z}\right)\left(\frac{\partial f_1}{\partial m} + \frac{\partial f_2}{\partial n}\right) = \left(\frac{\partial^2 f_1}{\partial m^2}\right) + \left(\frac{\partial^2 f_2}{\partial n^2}\right) \tag{17.14}$$

and

$$\frac{\partial H_x}{\partial t} = \left(\frac{\partial f_1}{\partial m}\right)\left(\frac{\partial m}{\partial t}\right) + \left(\frac{\partial f_2}{\partial n}\right)\left(\frac{\partial n}{\partial t}\right) = -\mathrm{u}\left(\frac{\partial f_1}{\partial m}\right) + \mathrm{u}\left(\frac{\partial f_2}{\partial n}\right) \tag{17.15}$$

and hence

$$\frac{\partial^2 H_x}{\partial t^2} = \left(\frac{\partial}{\partial m}\right)\left[-\mathrm{u}\left(\frac{\partial f_1}{\partial m}\right)\right]\left(\frac{\partial m}{\partial t}\right) + \left(\frac{\partial}{\partial n}\right)\left[\mathrm{u}\left(\frac{\partial f_2}{\partial n}\right)\right]\left(\frac{\partial n}{\partial t}\right)$$

$$= \mathrm{u}^2\left(\frac{\partial^2 f_1}{\partial m^2}\right) + \mathrm{u}^2\left(\frac{\partial^2 f_2}{\partial n^2}\right) = \mathrm{u}^2\left[\left(\frac{\partial^2 f_1}{\partial m^2}\right) + \left(\frac{\partial^2 f_2}{\partial n^2}\right)\right] \tag{17.16}$$

∴ From Eqs. (17.14) and (17.16), we obtain

$$\frac{\partial^2 H_x}{\partial z^2} = \left(\frac{1}{\mathrm{u}^2}\right)\left(\frac{\partial^2 H_x}{\partial t^2}\right) \tag{17.17}$$

which is same as Eq. (17.7) for H_x when

$$\mathrm{u} = \frac{1}{\sqrt{\mu\varepsilon}} \tag{17.18}$$

Thus each term, i.e.

$$f_1(z - \mathrm{u}t) \qquad \text{or} \qquad f_2(z + \mathrm{u}t) \tag{17.19}$$

is a solution of the **H** and **E** equation.

The sum of the two terms is a general solution.

Let us consider the first part of the solution, i.e.

$$H_{x:} = f_1(z - \mathrm{u}t) \tag{17.20}$$

which is the equation of a wave travelling in the +z-direction. The diagram (Figure17.1) shows $H_{x:}$ at $t = 0$ and $t = 1$, i.e. after 1 second, it will have the same distribution as it had at $t = 0$, i.e. the same solution in x- and y-coordinates, but it will be shifted by a distance u along the z-axis in the positive direction.

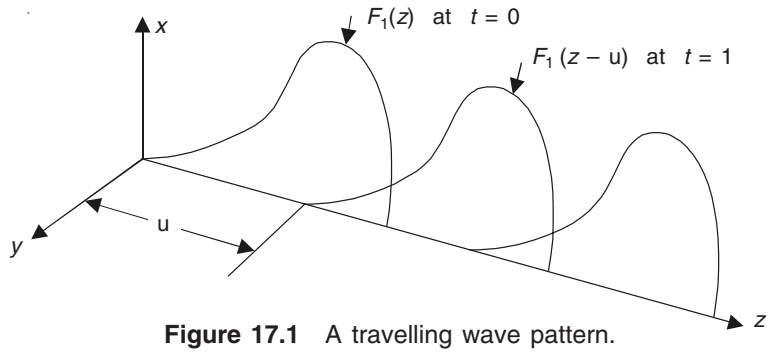


Figure 17.1 A travelling wave pattern.

Thus the wave $H_{x:} = f_1(z - \mathrm{u}t)$ moves a distance u along the z-axis in one second.

$\therefore$ Its velocity = $\mathrm{u} = \dfrac{1}{\sqrt{\mu_0\mu_\mathrm{r}\varepsilon_0\varepsilon_\mathrm{r}}}$ = velocity of light [in free space (= c)].

Similarly it can be proved that $H_{x:} = f_2(z + \mathrm{u}t)$ is a wave moving with velocity u in the (–z)-direction. In general, f_2 is not same as f_1, and hence the two waves need not necessarily have the same shape. In some cases, only one of these waves may be present, i.e. an isolated transmitting aerial. In other cases, the outgoing wave may be reflected back, thus producing two waves travelling in opposite directions at a point.

The most common type of wave is the 'sinusoidal wave', and examples of this type are: light waves, radio waves, etc. By solving the wave equation of the other components of **E** and **H**, we obtain relations of the same form. The **E** and **H** waves cannot be independent of one another because of the curl equations of Maxwell, i.e.

$$\nabla \times \mathbf{E} = -\mu_0\mu_\mathrm{r}\left(\frac{\partial \mathbf{H}}{\partial t}\right) \qquad \text{and} \qquad \nabla \times \mathrm{H} = \varepsilon_0\varepsilon_\mathrm{r}\left(\frac{\partial \mathbf{D}}{\partial t}\right)$$

Hence, to find the relationship between the components of **E** and **H**, we rewrite the equations for the scalar components of the field vectors from the above two equations; rewriting in the form:

$$\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = -\mu\left(\frac{\partial H_x}{\partial t}\right) \tag{17.21a}$$

$$\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} = -\mu\left(\frac{\partial H_y}{\partial t}\right) \tag{17.21b}$$

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = -\mu\left(\frac{\partial H_z}{\partial t}\right) \tag{17.21c}$$

$$\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} = \varepsilon\left(\frac{\partial E_x}{\partial t}\right) \tag{17.22a}$$

$$\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} = \varepsilon\left(\frac{\partial E_y}{\partial t}\right) \tag{17.22b}$$

$$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = \varepsilon\left(\frac{\partial E_z}{\partial t}\right) \tag{17.22c}$$

But for plane waves, we have defined $\partial\mathbf{E}/\partial x = 0 = \partial\mathbf{E}/\partial y$, i.e.

$$\frac{\partial E_z}{\partial x} = \frac{\partial E_z}{\partial y} = \frac{\partial E_y}{\partial x} = \frac{\partial E_y}{\partial y} = \frac{\partial E_x}{\partial x} = \frac{\partial E_x}{\partial y} = 0 \tag{17.23}$$

and similarly for **H**, i.e.

$$\frac{\partial H_z}{\partial x} = \frac{\partial H_z}{\partial y} = \frac{\partial H_y}{\partial x} = \frac{\partial H_y}{\partial y} = \frac{\partial H_x}{\partial x} = \frac{\partial H_x}{\partial y} = 0 \tag{17.24}$$

Substituting from Eqs. (17.23) and (17.24) in Eqs. (17.21) and (17.22), we obtain

$$H_z = 0 \qquad \text{and} \qquad E_z = 0 \tag{17.25}$$

and

$$\frac{\partial E_y}{\partial z} = \mu\left(\frac{\partial H_x}{\partial t}\right) \tag{17.26a}$$

$$\frac{\partial E_x}{\partial z} = -\mu\left(\frac{\partial H_y}{\partial t}\right) \tag{17.26b}$$

$$\frac{\partial H_y}{\partial z} = -\varepsilon\left(\frac{\partial E_x}{\partial t}\right) \tag{17.27a}$$

$$\frac{\partial H_x}{\partial z} = \varepsilon\left(\frac{\partial E_y}{\partial t}\right) \tag{17.27b}$$

But from Eq. (17.20), we have $H_{x:} = f_1(z - \mathrm{u}t)$.

$\therefore \quad \dfrac{\partial H_x}{\partial z} = f_1'(z - \mathrm{u}t)$, where f_1' is the derivative of f_1 with respect to $(z - \mathrm{u}t)$.

But

$$\frac{\partial H_x}{\partial z} = \varepsilon\left(\frac{\partial E_y}{\partial t}\right) \quad \text{and} \quad \therefore \ \frac{\partial E_y}{\partial t} = \left(\frac{1}{\varepsilon}\right) f_1'(z - \mathrm{u}t)$$

$\therefore$ Integrating with respect to time,

$$E_y = -\left(\frac{1}{\mathrm{u}\varepsilon}\right) f_1 (z - \mathrm{u}t) = -\left(\frac{1}{\mathrm{u}\varepsilon}\right) H_{x:}$$

But $\mathrm{u} = 1/\sqrt{\mu\varepsilon}$

$$\therefore \qquad E_y = -\sqrt{\frac{\mu}{\varepsilon}}\ H_{x:}, \text{ and similarly } E_x = \sqrt{\frac{\mu}{\varepsilon}}\ H_y \tag{17.28}$$

$$\therefore \quad \frac{|\mathbf{E}|}{|\mathbf{H}|} = \sqrt{\frac{\mu}{\varepsilon}} \tag{17.29}$$

where **E** and **H** are perpendicular to each other, and also perpendicular to the direction of propagation. The rotation from the electric field vector **E** to the magnetizing intensity vector **H** is right-handed about the direction of propagation. The ratio of $|\mathbf{E}|/|\mathbf{H}| \left(=\sqrt{\mu/\varepsilon}\right)$ at any point in space is known as the 'characteristic or wave impedance of the material'. In free space,

$$|\mathbf{E}| = \sqrt{\frac{\mu_0}{\varepsilon_0}}\ |\mathbf{H}| = Z_0\ |\mathbf{H}| \tag{17.30}$$

where Z_0 = characteristic impedance of free space. Its units: The unit of **E** is volts/metre, of **H** is amp-turn/metre, and hence the unit of Z is volts/amp or ohms (Ω).

In free space, the velocity

$$\mathrm{u}_0 = (\mu_0\varepsilon_0)^{-½} = \mathrm{c} = 2.998 \times 10^8 \text{ metres/sec} \tag{17.31}$$

i.e. the velocity of light.

The most common forms of electrical waves vary sinusoidally with time and position. If the frequency is = f Hz, and the wavelength = λ, then the frequency of repetition of the wave pattern is

$$f = \frac{\mathrm{u}}{\lambda} \qquad \text{or} \qquad f\lambda = \mathrm{u} \tag{17.32}$$

The angular frequency ω is

$$\omega = 2\pi f = \frac{2\pi\mathrm{u}}{\lambda} \tag{17.33}$$

$$\therefore \qquad \frac{\omega}{\mathrm{u}} = \frac{2\pi}{\lambda} = \beta \text{ (or sometimes } k) \tag{17.34}$$

where β is the wave number of the wave.

So we can now write the general expressions for both the electric and the associated magnetic wave, i.e.

$$E_x = E_{\mathrm{F}x} \sin\left[\left(\frac{2\pi}{\lambda}\right)(\mathrm{u}t - z)\right] + E_{\mathrm{B}x} \sin\left[\left(\frac{2\pi}{\lambda}\right)(\mathrm{u}t + z)\right]$$

$$= E_{Fx} \sin\left[\omega\left(t - \frac{z}{u}\right)\right] + E_{Bx} \sin\left[\omega\left(t + \frac{z}{u}\right)\right]$$

$$= E_{Fx} \sin(\omega t - \beta z) + E_{Bx} \sin(\omega t + \beta z) \tag{17.35}$$

Similarly,

$$H_x = H_{Fx} \sin(\omega t - \beta z) + H_{Bx} \sin(\omega t + \beta z) \tag{17.36}$$

where

E_{Fx} = amplitude of the x-component of the forward travelling **E** wave

E_{Bx} = amplitude of the x-component of the backward travelling **E** wave

Similarly for the **H** wave.

Usually the exponential equation and the complex j notation are used, i.e.

$$E_x = E_{Fx} \exp[j(\omega t - \beta z)] + E_{Bx} \exp[j(\omega t + \beta z)] \tag{17.37a}$$

and

$$H_x = H_{Fx} \exp[j(\omega t - \beta z)] + H_{Bx} \exp[j(\omega t + \beta z)] \tag{17.37b}$$

It should be noted that the forward and the backward travelling waves involve a sign change in their corresponding characteristic equations, i.e.

$$E_{Fx} = \sqrt{\frac{\mu}{\varepsilon}}\, H_{Fy}, \quad E_{Fy} = -\sqrt{\frac{\mu}{\varepsilon}}\, H_{Fx:} \tag{17.38a}$$

and

$$E_{Bx} = -\sqrt{\frac{\mu}{\varepsilon}}\, H_{By}, \quad E_{By} = \sqrt{\frac{\mu}{\varepsilon}}\, H_{Bx:} \tag{17.38b}$$

∴ The relations of Eqs. (17.28) and (17.29) also hold for both the forward-going and the backward-going waves, i.e.

$$\frac{E_{Fx}}{H_{Fy}} = -\frac{E_{Fx}}{H_{Fx}} = \sqrt{\frac{\mu}{\varepsilon}} = Z \tag{17.39a}$$

$$-\frac{E_{Bx}}{H_{By}} = \frac{E_{By}}{H_{Bx}} = \sqrt{\frac{\mu}{\varepsilon}} = Z \tag{17.39b}$$

i.e. **E** and **H** are perpendicular to one another in each travelling wave, and the ratio of E to H is same in both the waves.

The vector (**E** × **H**), i.e. the Poynting vector for the wave follows the direction of propagation of the wave in each case.

So next, we consider the energy relations for the electromagnetic wave.

∴ Stored energy in the electric field/unit volume = W_e

$$= \frac{\varepsilon |\mathbf{E}|^2}{2} = \left(\frac{\varepsilon}{2}\right)\left(E_x^2 + E_y^2\right) \tag{17.40}$$

and, the stored energy in the magnetic field/unit volume = W_m

$$= \frac{\mu |\mathbf{H}|^2}{2} = \left(\frac{\mu}{2}\right)\left(H_x^2 + H_y^2\right) \tag{17.41}$$

But, Eq. (17.40) = Eq. (17.41), i.e.

$$\left(\frac{\varepsilon}{2}\right)\left(E_x^2 + E_y^2\right) = \left(\frac{\mu}{2}\right)\left(H_x^2 + H_y^2\right) \tag{17.42}$$

$\therefore$ The energy density at each point, at each instant, is equally divided between the electric and the magnetic energy.

Poynting vector for the forward wave = $E_{\mathrm{F}x}H_{\mathrm{F}y} - E_{\mathrm{F}y}H_{\mathrm{F}x}$

$$= \left(\frac{1}{Z}\right)\left(E_{\mathrm{F}x}^2 + E_{\mathrm{F}y}^2\right) = Z\left(H_{\mathrm{F}x}^2 + H_{\mathrm{F}y}^2\right) \tag{17.43}$$

Similarly for the backward travelling wave,

the corresponding Poynting vector = $E_{\mathrm{B}x}H_{\mathrm{B}y} - E_{\mathrm{B}y}H_{\mathrm{B}x}$

$$= \left(\frac{1}{Z}\right)\left(E_{\mathrm{B}x}^2 + E_{\mathrm{B}y}^2\right) = Z\left(H_{\mathrm{B}x}^2 + H_{\mathrm{B}y}^2\right) \tag{17.44}$$

When a forward travelling wave and a backward travelling wave are of equal amplitude and of same phase, the combination produces a pattern which is stationary in space and does not travel with time; i.e. from Eqs. (17.35), if $E_{\mathrm{B}x} = E_{\mathrm{F}x}$, then we have

$$\begin{aligned} E_x &= E_{\mathrm{F}x}[\sin(\omega t - \beta z) + \sin(\omega t + \beta z)] \\ &= 2E_{\mathrm{F}x}[\sin(\beta z)\cos(\omega t)] \end{aligned} \tag{17.45}$$

This is a sinusoidal wave fixed in space, with its amplitude fluctuating at the frequency $f\,(= \omega/2\pi)$.

For $\beta z = (2\pi/\lambda)z$, $\sin(\beta z) = \sin[(2\pi/\lambda)z]$. This will be zero, when z = a multiple of $(\lambda/2)$.

$\therefore$ $E_x = 0$ at these points for all instants of time.

These are the *NODES* in the electric field (Figure 17.2).

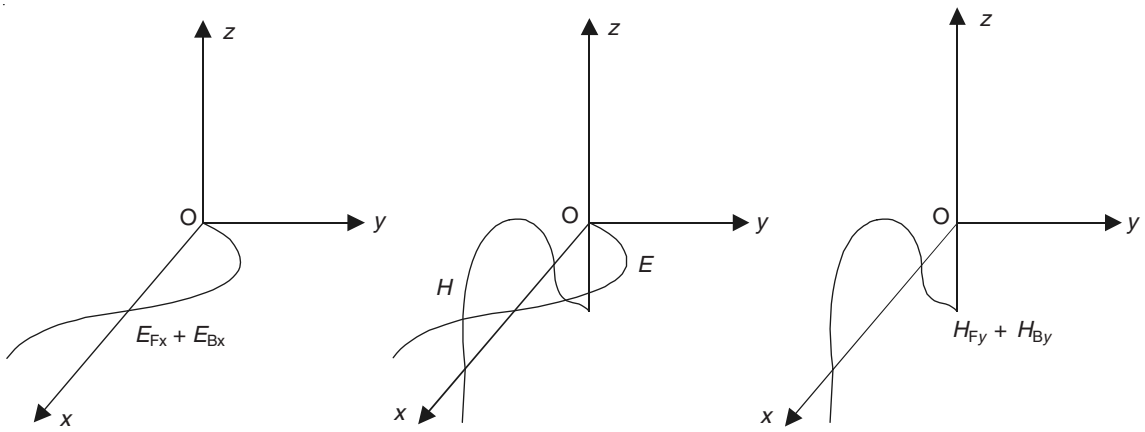


Figure 17.2 Standing wave patterns.

The associated magnetic field is H_y or B_y, and hence the corresponding standing magnetic field (wave) will be

$$H_y = -2H_{Fy} [\cos(\beta z) \sin(\omega t)] \tag{17.46}$$

Its nodes are half-way through the nodes of the **E** field. So we see that the standing waves are produced by the mutual interference of two waves of the same periodicity.

17.2.1 Some General Comments on Uniform Plane Waves

Maxwell's equations do not impose any limits on the frequency of the electromagnetic waves. Uptill now the spectrum that has been studied experimentally, ranges from the long radio waves (at frequencies of about 10^4 Hz and wavelengths about 3×10^4 metres) to the very high energy gamma rays (of frequencies 10^{24} Hz and higher, and of wavelengths of the order of 3×10^{-16} metres and shorter) noticed in cosmic radiation. The known spectrum thus covers a range of greater than 20 orders of magnitude. Radio, light, heat waves, X-rays, and gamma rays are all electromagnetic, though their sources, detectors as well as their modes of interaction with matter, vary widely as the frequency changes. A fundamental identity of all these types of waves is demonstrated by the fact that in free space, they are all transverse waves with a common velocity of propagation c (i.e. the velocity of light).

We have also seen from our analysis in Section 12.2 that (for example) a plane electromagnetic wave propagating in the positive z-direction, **E** is independent of x and y [Eq. (17.5)]

$$\therefore \qquad \nabla \cdot \mathbf{E} = \frac{\partial E_z}{\partial z} = 0 \tag{17.47}$$

i.e. the z-component of **E** cannot be a function of z, and this was found to be equal to zero [i.e. Eq. (17.5a)]; and the same argument applies to the **H** field. (Note that at present, our interest is in the uniform plane 'wave' and *not* in the 'uniform field'.)

Thus a plane electromagnetic wave propagating in free space is therefore 'transverse', since it has no longitudinal components. A uniform plane wave is also called a 'plane-polarized wave'. So now we define the term 'plane of polarization of a wave'.

The plane of polarization is the plane containing the direction of propagation and the **E** vector (though originally it used to be the **H** vector). A plane polarized wave is also called 'linearly polarized'. Since the **E** vector is the reference vector, 'vertical polarization' would mean that the **E** vector is in the vertical direction, and the 'horizontal polarization' means that the **E** vector is in the horizontal direction (Figure 17.3). Also, without any loss of generality, we can assume a wave to be plane polarized with its **E** vector along the direction of the x-axis. This is

Figure 17.3 (a) Vertically polarized and (b) horizontally polarized wave representation.

because any plane-polarized wave can be considered to be the sum of two waves that are plane-polarized in perpendicular directions, and in phase. This is shown in Figure 17.4.

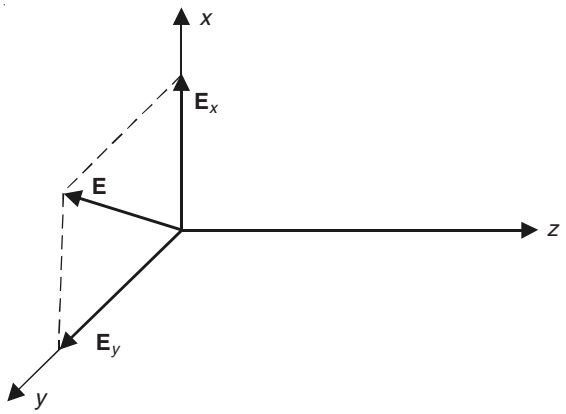


Figure 17.4 Resolution of **E** into two orthogonal components.

We can also add two plane-polarized waves (orthogonal space wise) that differ in phase, i.e.

$$E_x = E_1 \cos \omega\left(t - \frac{z}{\mathrm{u}}\right) \quad \text{and} \quad E_y = E_2 \cos\left[\omega\left(t - \frac{z}{\mathrm{u}}\right) + \psi\right] \tag{17.48}$$

For these two waves at $z = 0$,

$$E_x = E_1 \cos \omega t \quad \text{and} \quad E_y = E_2 \cos (\omega t + \psi) \tag{17.49}$$

The equations of (17.49) are the parametric equations of an ellipse. For this combination, the maxima of $\mathbf{E}_x$ and $\mathbf{E}_y$ do not occur at the same time, and their sum **E** describes an ellipse about the z-axis. Such a resultant wave is called an 'elliptically polarized wave'.

If E_x and E_y have equal amplitudes, but are 90° out of phase, i.e. $E_1 = E_2$, and $\psi = \pi/2$, then:

$$E_x^2 + E_y^2 = E_1^2 \tag{17.50}$$

i.e. the ellipse has degenerated into a circle, and the wave is said to be circularly polarized.

The polarization is said to be right-handed or left-handed depending on whether the vectors **E** and **H** rotate clockwise or counterclockwise for an observer looking at the source.

In discussing the propagation of waves, we use the term wave-front, which we shall explain formally. When a wave is progressing, let us consider a surface S, separating region 1 (in Figure 17.5) which has been affected by the disturbance, from the region 2 which has *not* been affected. In this, we neglect any superimposed magnetic field, say, like the earth's magnetic field. When we say that the velocity of propagation of the wave is c, we mean that the position of the surface S, after a further time δt has shifted to a position S' which lies a distance $\mathrm{c}\delta t$ along the normal to S in the region which was originally region 2 and this region covered by $\mathrm{c}\delta t$ now lies in the region 1. So when we say that the plane wave is a transverse wave with the vectors **E** and **H** at right angles to the direction of propagation, it implies that **E** and **H** are coplanar with the wave-front surface S.

So far, when we talked about the electromagnetic waves, we have been referring to **E** and **H** waves, which are related by the equation

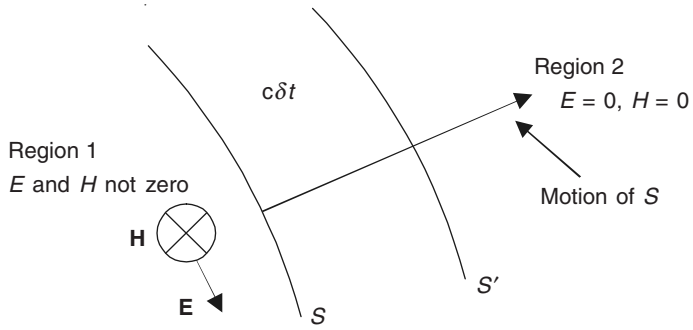


Figure 17.5 Wave-front progressing.

$$|\mathbf{E}| = Z_0 |\mathbf{H}|$$

where $Z_0 = 377\ \Omega = \sqrt{\mu_0/\varepsilon_0}$, and is the characteristic impedance of free space. However, if instead of referring to the **H** as the magnetic wave, we were to use the **B**, then the relationship between **E** and **B** would be $\mathbf{B} = \mu_0\mathbf{H}$ in free space,

$$|\mathbf{E}| = \left(\frac{Z_0}{\mu_0}\right)|\mathbf{H}| = \mathrm{c}|\mathbf{B}| \tag{17.51}$$

17.3 THE E AND H VECTORS IN LINEAR, ISOTROPIC, HOMOGENEOUS AND STATIONARY MEDIA (LIH MEDIA)

We start by defining these terms:

A medium is homogeneous if its properties do not vary from point to point. A medium is isotropic if its properties are same in all directions from any given point. A medium is linear and isotropic if

$$\mathbf{D} = \varepsilon\mathbf{E}, \quad \mathbf{H} = \frac{\mathbf{B}}{\mu}, \quad \mathbf{J} = \sigma\mathbf{E} \tag{17.52}$$

where ε, μ, and σ are constants independent of **E** and **H**, and independent of direction.

The wave equations for such media follow from Eqs. (12.32) and (12.35), which give us

$$\nabla^2\mathbf{E} - \mu\sigma\left(\frac{\partial\mathbf{E}}{\partial t}\right) - \mu\varepsilon\left(\frac{\partial^2\mathbf{E}}{\partial t^2}\right) = 0 \tag{12.35}$$

and

$$\nabla^2\mathbf{H} - \mu\sigma\left(\frac{\partial\mathbf{H}}{\partial t}\right) - \mu\varepsilon\left(\frac{\partial^2\mathbf{H}}{\partial t^2}\right) = 0 \tag{12.32}$$

assuming charge-free region.

We also assume that the plane wave is propagating in positive direction of the z-axis. Hence $(\partial/\partial x) = 0$ and $(\partial/\partial y) = 0$. Then the equation for the divergence of **D** becomes:

$$\nabla \cdot \mathbf{D} = \nabla \cdot (\varepsilon\mathbf{E}) = \varepsilon\left(\frac{\partial E_z}{\partial z}\right) = 0 \text{ (no free charge)} \tag{17.53}$$

$$\therefore \quad \frac{\partial^2 E_z}{\partial z^2} = 0 \tag{17.54}$$

Then the wave equation for **E** [i.e. Eq. (12.35)] becomes:

$$\left(\frac{\partial^2}{\partial z^2}\right)(\mathbf{i}_x E_x + \mathbf{i}_y E_y) - \mu\left(\sigma\frac{\partial}{\partial t} + \varepsilon\frac{\partial^2}{\partial t^2}\right)(\mathbf{i}_x E_x + \mathbf{i}_y E_y + \mathbf{i}_z E_z) = 0 \tag{12.35}$$

$\therefore$ The longitudinal component of **E**, i.e. E_z would satisfy the equation

$$\varepsilon\left(\frac{\partial^2 E_z}{\partial t^2}\right) + \sigma\left(\frac{\partial E_z}{\partial t}\right) = 0 \tag{17.55}$$

So, if E_z exists, it will be of the form:

$$E_z = a + b \ \exp\left(\frac{-\sigma t}{\varepsilon}\right) \tag{17.56}$$

a and b being the constants of integration [a form which is consistent with the requirements of Eq. (17.53)]. Thus E_z would decrease exponentially with time, and there is no E_z wave. If $\sigma = 0$, then E_z would be of the form $a + bt$, and again there is no E_z wave. Hence, for waves, we can say: $E_z = 0$.

Note: We would come to the same conclusion even if the charge density ρ_C was not zero.

Thus, for plane electromagnetic waves, the **E** vector is transverse in LIH media. We can similarly prove that the **H** vector is also a transverse wave.

Hence, the plane electromagnetic waves are transverse in any LIH and stationary medium. We shall now have a look at their relative orientation. Without any loss of generality, we can assume a plane-polarized wave with the **E** vector in the x–z plane, i.e. parallel to the x-axis, and sinusoidally varying with time. Hence

$$\mathbf{E} = \mathbf{i}_x E_{Ox} \exp\,[j(\omega t - kz)] \tag{17.57}$$

where k is in general complex. Then from the Maxwell's equation

$$\nabla \times \mathbf{E} = -\left(\frac{\partial \mathbf{B}}{\partial t}\right)$$

we get

$$\mathbf{i}_x 0 + \mathbf{i}_y(-jkE_{Ox}) + \mathbf{i}_z 0 = -j\omega\mu\,(\mathbf{i}_x H_x + \mathbf{i}_y H_y)$$

$$\therefore \qquad H_x = 0 \qquad \text{and} \qquad \mathbf{H} = \mathbf{i}_y H_y = \mathbf{i}_y\left(\frac{k}{\omega\mu}\right)E_{Ox} \exp\,[j(\omega t - kz)] \tag{17.58}$$

$$\therefore \qquad \left|\frac{\mathbf{E}}{\mathbf{H}}\right| = \frac{\omega\mu}{k} \tag{17.59}$$

Hence the **E** and **H** vectors in a plane-polarized wave in a LIH medium are (1) orthogonal (i.e. mutually perpendicular), (2) so oriented that their vector product ($\mathbf{E} \times \mathbf{H}$) points in the direction of propagation, and (3) not necessarily in phase because the wave number k can be complex, and hence the waves can get attenuated as they propagate.

17.4 PROPAGATION OF PLANE ELECTROMAGNETIC WAVES IN PERFECT LOSS-LESS NON-CONDUCTORS (INSULATORS OR DIELECTRICS)

In loss-less dielectrics, $\sigma = 0$, and as a rule, $\rho_C = 0$. Hence the wave equations (12.35) and (12.32) reduce to

$$\nabla^2 \mathbf{E} - \mu\varepsilon\left(\frac{\partial^2 \mathbf{E}}{\partial t^2}\right) = 0 \tag{17.60a}$$

and

$$\nabla^2 \mathbf{H} - \mu\varepsilon\left(\frac{\partial^2 \mathbf{H}}{\partial t^2}\right) = 0 \tag{17.60b}$$

E and **H** vectors are then of the same form as in Section 17.3, i.e.

$$\mathbf{E} = \mathbf{i}_x E_{Ox} \exp\,[\,j(\omega t - \beta z)] \tag{17.61}$$

Since $\sigma = 0$, we shall see that the wave number β will not be complex. From Eq. (17.61), the **H** wave comes out to be (as in Section 17.3):

$$\mathbf{H} = \mathbf{i}_y\left(\frac{\beta}{\omega\mu}\right) E_{Ox} \exp\,[\,j(\omega t - \beta z)] \tag{17.62}$$

Substituting for **E** and **H** from Eqs. (17.61) and (17.62) in Eq. (17.60), we find

$$-\beta^2 + \omega^2\mu\varepsilon = 0 \tag{17.63}$$

$\therefore$ The wave number, $$\beta = \omega\sqrt{\mu\varepsilon} = \frac{\omega}{\mathrm{u}} \tag{17.64}$$

is real, and hence there is no attenuation of either of the two waves. The phase velocity

$$\mathrm{u} = \frac{\omega}{\beta} = \frac{1}{\sqrt{\mu\varepsilon}} = \frac{\mathrm{c}}{\sqrt{\mu_r \varepsilon_r}} \tag{17.65}$$

$\therefore$ The phase velocity of the waves in the non-conductors is less than that in free space, and the 'index of refraction', n is

$$n = \frac{\mathrm{c}}{\mathrm{u}} = \sqrt{\mu_r \varepsilon_r} \tag{17.66}$$

In a non-magnetic medium, $\mu_r = 1$, and

$$n = \sqrt{\varepsilon_r} \tag{17.67}$$

Note that, in general, n and ε_r are both functions of frequency.

In the non-conductors, the **E** and **H** vectors are in phase, and the electric and the magnetic energy densities are equal, i.e.

$$\frac{1}{2}\varepsilon E^2 = \frac{1}{2}\mu H^2$$

$\therefore$ The total instantaneous energy density = $\varepsilon E^2 = \mu H^2$

The average total energy density = $\varepsilon E_{rms}^2 = \mu H_{rms}^2$

The average value of the Poynting vector:

$$\bar{\mathbf{S}} = \mathbf{i}_z \left(\frac{1}{2}\right) E_{Ox} H_{Ox} = \mathbf{i}_z \left(\frac{1}{2}\right) \sqrt{\frac{\varepsilon}{\mu}}\, E_{Ox}^2$$

$$= \mathbf{i}_z \sqrt{\frac{\varepsilon}{\mu}}\, E_{\text{rms}}^2 = \mathbf{i}_z (\text{u}\varepsilon)\, E_{\text{rms}}^2 \text{ watts/metre}^2 \tag{17.68}$$

The average value of the Poynting vector is thus equal to the phase velocity u multiplied by the average energy density.

17.5 PROPAGATION OF PLANE ELECTROMAGNETIC WAVES IN CONDUCTING MEDIA

For a conducting medium $\sigma \neq 0$, the wave equations (12.35) and (12.32) for a wave travelling along the z-axis, with $\rho_C = 0$, have to be solved, i.e.

$$\nabla^2 \mathbf{E} - \mu\sigma\left(\frac{\partial \mathbf{E}}{\partial t}\right) - \mu\varepsilon\left(\frac{\partial^2 \mathbf{E}}{\partial t^2}\right) = 0 \tag{12.35}$$

$$\nabla^2 \mathbf{H} - \mu\sigma\left(\frac{\partial \mathbf{H}}{\partial t}\right) - \mu\varepsilon\left(\frac{\partial^2 \mathbf{H}}{\partial t^2}\right) = 0 \tag{12.32}$$

The solutions of **E** and **H** vectors with sinusoidal time variations come out to be:

$$\mathbf{E} = \mathbf{i}_x E_{Ox} \exp\,[\,j(\omega t - kz)] \tag{17.69}$$

$$\mathbf{H} = \mathbf{i}_y \left(\frac{k}{\omega\mu}\right) E_{Ox} \exp\,[\,j\,(\omega t - kz)] = \mathbf{i}_y H_{Oy} \exp\,[\,j(\omega t - kz)] \tag{17.70}$$

By substituting, we get

$$-k^2 + \omega^2\mu\varepsilon - j\omega\mu\sigma = 0$$

or

$$k^2 = \omega^2\mu\varepsilon - j\omega\mu\sigma = \omega^2\mu\varepsilon\left[1 - j\left(\frac{\sigma}{\omega\varepsilon}\right)\right]$$

$$= \omega^2\mu_0\mu_r\varepsilon_0\varepsilon_r\left[1 - j\left(\frac{\sigma}{\omega\varepsilon}\right)\right] = \left(\frac{\omega^2\mu_r\varepsilon_r}{c^2}\right)\left[1 - j\left(\frac{\sigma}{\omega\varepsilon}\right)\right]$$

$$= \left(\frac{4\pi^2\mu_r\varepsilon_r}{\lambda_0^2}\right)\left[1 - j\left(\frac{\sigma}{\omega\varepsilon}\right)\right], \qquad \frac{\omega}{c} = \frac{2\pi}{\lambda_0} \tag{17.71}$$

where λ_0 is the wavelength in free space.

As k^2 is complex, we write

$$k = k_r - jk_i \tag{17.72}$$

$$\therefore \qquad k_r = \left(\frac{2\pi}{\lambda_0}\right)\sqrt{\frac{\mu_r\varepsilon_r}{2}}\left[\left(1 + \frac{\sigma^2}{\omega^2\varepsilon^2}\right)^{1/2} + 1\right]^{1/2}$$

and

$$k_i = \left(\frac{2\pi}{\lambda_0}\right)\sqrt{\frac{\mu_r \varepsilon_r}{2}}\left[\left(1+\frac{\sigma^2}{\omega^2\varepsilon^2}\right)^{1/2}-1\right]^{1/2} \tag{17.73}$$

The real part of the wave number k, i.e. k_r is $(2\pi/\lambda)$, where λ is the wavelength in the medium. The imaginary part k_i is the reciprocal of the distance δ over which the amplitude is attenuated by a factor of e. The quantity $\delta = 1/k_i$ is called the 'attenuation distance'.

The phase velocity is

$$\mathrm{u} = \frac{\omega}{k_r} \tag{17.74}$$

corresponding to an index of refraction

$$n = \frac{\mathrm{c}}{\mathrm{u}} = \left(\frac{\mathrm{c}}{\omega}\right)k_r = \left(\frac{\lambda_0}{2\pi}\right)k_r = \frac{\lambda_0}{\lambda} \tag{17.75}$$

and following Eq. (17.59):

$$\left|\frac{\mathbf{E}}{\mathbf{H}}\right| = \frac{\omega\mu}{k}$$

$$= \sqrt{\frac{\mu}{\varepsilon}}\left[\left(1+\frac{\sigma^2}{\omega^2\varepsilon^2}\right)\right]^{-1/4} \exp\left[j\tan^{-1}\left(\frac{k_i}{k_r}\right)\right] \tag{17.76}$$

the quantity $[\tan^{-1}(k_i/k_r)]$ denotes the phase of $\mathbf{E}$ with respect to $\mathbf{H}$.

The expressions for $\mathbf{E}$ and $\mathbf{H}$ vectors are

$$\mathbf{E} = \mathbf{i}_x E_{\mathrm{O}x} \exp\left[j(\omega t - k_r z) - k_i z\right]$$

and

$$\mathbf{H} = \mathbf{i}_y H_{\mathrm{O}y} \exp\left[j\left(\omega t - k_r z - \tan^{-1}\left(\frac{k_i}{k_r}\right) - k_i z\right)\right] \tag{17.77}$$

where

$$\frac{E_{\mathrm{O}x}}{H_{\mathrm{O}y}} = \sqrt{\frac{\mu}{\varepsilon}}\left(1+\frac{\sigma^2}{\omega^2\varepsilon^2}\right)^{-1/4} \tag{17.78}$$

The ratio of the electric energy to the magnetic energy is

$$\left|\frac{(1/2)\,\varepsilon E^2}{(1/2)\,\mu H^2}\right| = \left(1+\frac{\sigma^2}{\omega^2\varepsilon^2}\right)^{-1/2} \tag{17.79}$$

17.6 PROPAGATION OF PLANE ELECTROMAGNETIC WAVES IN GOOD CONDUCTORS

In Section 17.5, we studied the propagation of electromagnetic waves in conductors. So now to define what we mean by 'good' conductors, we start at the values of k_r and k_i for conducting media as obtained in Eqs. (17.73). In good conductors, the ratio $\omega\varepsilon/\sigma$ [as we have already mentioned in Chapter 12, Section 12.5, Eqs. (12.37) and (12.39)] is much smaller than unity, i.e.

$$\frac{\omega\varepsilon}{\sigma} \leq \frac{1}{50} \text{ (say)} \tag{17.80}$$

which means that the conduction current density ($\mathbf{J} = \sigma\mathbf{E}$) must be at least 50 times greater than the displacement current density ($\partial\mathbf{D}/\partial t$). So we define 'good conductors' as those for which the above condition is satisfied. Hence copper is a good conductor for frequencies up to about 2×10^{16} Hz or to the ultraviolet.

For good conductors, Eq. (17.71) simplifies to

$$k^2 = -j\omega\mu\sigma \tag{17.81}$$

and

$$\therefore \quad k = \sqrt{-j\omega\mu\sigma} = \left(\sqrt{\frac{\omega\mu\sigma}{2}}\right)(1 - j) \tag{17.82}$$

Hence, now

$$k_r = k_i = \sqrt{\frac{\omega\mu\sigma}{2}}$$

$$\therefore \qquad \delta = d\sqrt{2} = \sqrt{\frac{2}{\omega\mu\sigma}} = \frac{\lambda}{2\pi} \tag{17.83}$$

And, from Eq. (17.59), for the plane-polarized waves

$$\left|\frac{\mathbf{E}}{\mathbf{H}}\right| = \frac{\omega\mu}{k} = \sqrt{\frac{\omega\mu}{\sigma}} \exp\left(\frac{j\pi}{4}\right) \tag{17.84}$$

$\therefore$ In good conductors, the **E** vector leads the **H** vector by $\pi/4$, whereas for the non-conductors (i.e. dielectrics) we found that **E** and **H** are in phase. The reason for this difference is due to the fact that for good conductors, the **H** vector is from the conduction current, and for loss-less dielectrics, the **H** vector is produced by the displacement currents. Hence we can write the expressions for **E** and **H** for good conductors, similar to Eqs. (17.69) and (17.70), i.e.

$$\mathbf{E} = \mathbf{i}_x E_{Ox} \exp\left[j\left(\omega t - \frac{z}{d\sqrt{2}}\right) - \frac{z}{d\sqrt{2}}\right] \tag{17.85a}$$

$$\mathbf{H} = \mathbf{i}_y \left(\frac{\sigma}{\omega\mu}\right)^{1/2} E_{Ox} \exp\left[j\left(\omega t - \frac{z}{d\sqrt{2}} - \frac{\pi}{4}\right) - \frac{z}{d\sqrt{2}}\right] \tag{17.85b}$$

or, if we write in terms of cosine functions:

$$\mathbf{E} = \mathbf{i}_x E_{Ox} \exp\left(-\frac{z}{d\sqrt{2}}\right)\cos\left(\omega t - \frac{z}{d\sqrt{2}}\right) \tag{17.86a}$$

$$\mathbf{H} = \mathbf{i}_y \left(\frac{\sigma}{\omega\mu}\right)^{1/2} E_{Ox} \exp\left(-\frac{z}{d\sqrt{2}}\right)\cos\left(\omega t - \frac{z}{d\sqrt{2}} - \frac{\pi}{4}\right)$$

$$= \mathbf{i}_y H_{Oy} \exp\left(-\frac{z}{d\sqrt{2}}\right)\cos\left(\omega t - \frac{z}{d\sqrt{2}} - \frac{\pi}{4}\right) \tag{17.86b}$$

These results are similar to those, which we have obtained for good conductors in Chapter 15, Section 15.2. The distributions for **E** and **H** would be same as discussed there. The quantity δ derived here is the 'skin depth ($d\sqrt{2}$)' of the above-mentioned section.

The amplitude of the wave is attenuated by a factor of $(1/e) = 0.368$ in one radian length $(\lambda/2\pi)$, and by a factor of $(1/e)^{2\pi} \simeq 2 \times 10^{-3}$ in one wavelength λ, whereas the Poynting vector $(\mathbf{E} \times \mathbf{H})$ is attenuated by $(1/e)^2 = 0.135$ in $(\lambda/2\pi)$, and by $(1/e)^{4\pi} \simeq 4 \times 10^{-6}$ in one wavelength λ. The attenuation is so rapid that the wave is barely noticeable. Table 17.1 shows the skin-depth for various conductors at different frequencies. It should be noted that the attenuation in iron is much larger compared with that in silver, even though iron is a poor conductor compared with silver because of the fact that iron's permeability is much higher.

Table 17.1 Skin-depth for conductors (mm)

Conductor	σ (mho/m)	μ_r	*Skin-depth* [$(d\sqrt{2})$ mm]			
			60 Hz	500 Hz	10^4 Hz	10^6 Hz
Silver	6.15×10^7	1.00	8.48	2.94	0.66	0.066
Copper	5.8×10^7	1.00	8.63	2.99	0.68	0.068
Aluminium	3.54×10^7	1.00	10.1	3.81	0.85	0.085
Iron	1.0×10^7	1000	0.65	0.22	0.05	0.005

The phase velocity

$$\mathrm{u} = \frac{\omega}{k_r} = \frac{\omega\lambda}{2\pi} = \sqrt{\frac{2\omega}{\mu\sigma}} \tag{17.87}$$

is proportional to the square root of the frequency.

The ratio of the electric to the magnetic energy density is

$$\frac{(1/2)\,\varepsilon E^2}{(1/2)\,\mu H^2} = \frac{\omega\varepsilon}{\sigma} \leq \frac{1}{50} \tag{17.88}$$

and so most of the energy is in the magnetic form. This is a consequence of the medium being 'good' conductor which causes E/J to be small. Hence the electric field intensity is weak, but the current density, and hence **H** is relatively larger.

17.7 REFLECTION AND REFRACTION OF PLANE ELECTROMAGNETIC WAVES

So far we have studied the propagation of the electromagnetic waves in different types of infinite continuous media. We shall now study the effects of discontinuity in the medium of propagation. We shall investigate again the behaviour of these waves in different combinations of different types of media, i.e. dielectrics, conductors (including good conductors) and combinations of these. We remind ourselves that the dielectrics are non-conductors, and they may be either magnetic or non-magnetic; we shall restrict ourselves to non-magnetic dielectrics.

We assume an ideally thin, infinitely plane interface between the two linear, isotropic and homogeneous (LIH) media. An incident wave along $\mathbf{n}_i$ would, in general, give rise to a reflected wave along $\mathbf{n}_r$, and a transmitted wave along $\mathbf{n}_t$. These three waves, combined together, satisfy

the continuity conditions for the tangential components of **E** and **H**, and for the normal components of **D** and **B** at the interface (Figure 17.6). During the initial stages of our study, we shall exclude the total reflection from the dielectric, and we assume the media to extend to infinity on both the sides of the interface.

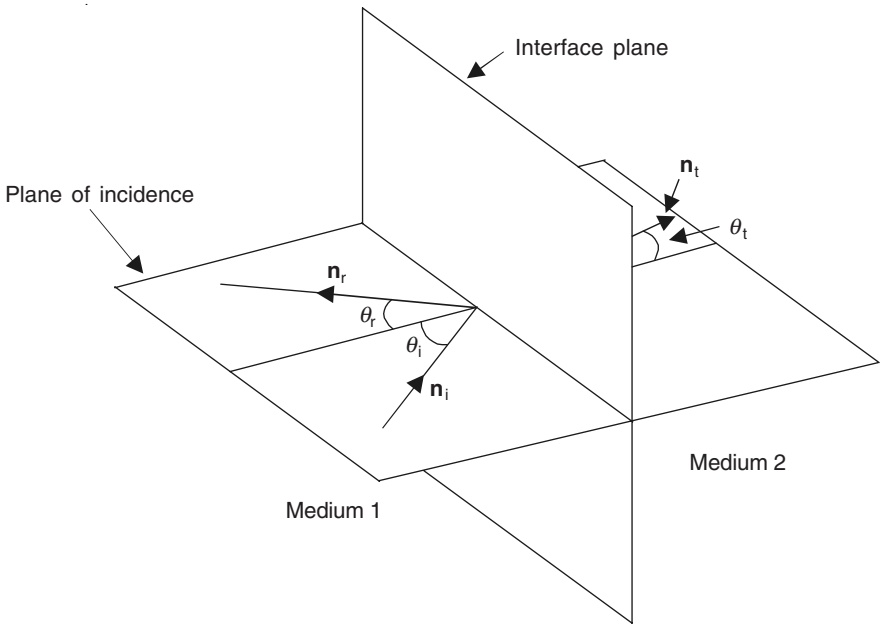


Figure 17.6 An electromagnetic wave in medium 1, incident on the interface between media 1 and 2, giving rise to both the reflected wave in the medium 1 and the transmitted wave in the medium 2.

For the present study we shall classify our analysis not by the different types of media discontinuities but by the way by which the incident wave hits the interface surface of discontinuity, i.e. (1) the incident wave meeting the interface normally; and (2) the incident wave meeting the interface obliquely, in which case, the interference pattern produced in the first medium, by the combination of the incident and the reflected waves would be stationary in one direction and travelling in an orthogonal direction. The properties of such patterns are of great practical importance in design and study of waveguides.

So before we start our study of reflection and refraction of waves, we shall remind ourselves of the facts that when a wave meets a discontinuity in the media of propagation, then, in general, it (the incident wave) produces a transmitted wave in the second medium and a reflected wave which travels back in the first medium, the exact directions of propagation being decided upon by the interface continuity conditions. We shall now consider waves incident normally on different types of interface surfaces, starting first with a perfectly conducting surface. However, before we start discussing each case separately, we shall state, in general terms, some fundamental points about the laws of electromagnetic reflection and refraction, and some basic points regarding the techniques of solving such problems.

17.8 THE LAWS OF REFLECTION AND SNELL'S LAW OF REFRACTION

Referring to Figure 17.6, we consider the electromagnetic wave incident on the interface, to be both plane and plane-polarized, so that its electric field intensity is of the form

$$\mathbf{E}_{1i} = \mathbf{E}_{0i} \exp\,[j\omega_i(t - \mathbf{n}_i \cdot \mathbf{r}/u_1)] \tag{17.89}$$

where u_1 is the phase velocity of the wave in the medium 1. The time $t = 0$, and the origin $r = 0$ can be chosen arbitrarily. We, however, choose the origin at a convenient point on the interface plane. The above equation defines a plane wave for all values of t and $\mathbf{r}$, but we shall use it only for medium 1, and when we consider normal incidence, then $\mathbf{n}_i \cdot \mathbf{r} = z$ (+ve z-direction). The reflected and the refracted waves from the plane interface are also plane and plane-polarized (obeying the same laws at all points on the interface), and are of the form

$$\mathbf{E}_r = \mathbf{E}_{0r} \exp\,[j\omega_r(t - \mathbf{n}_r \cdot \mathbf{r}/u_1)] \tag{17.90a}$$

and

$$\mathbf{E}_t = \mathbf{E}_{0t} \exp\,[j\omega_t(t - \mathbf{n}_t \cdot \mathbf{r}/u_2)] \tag{17.90b}$$

where u_2 is the phase velocity of the wave in the medium 2. It should be noted that so far no assumptions have been made about the amplitudes, phases, frequencies, and the directions of the reflected and the refracted waves. The amplitudes $\mathbf{E}_{0r}$, $\mathbf{E}_{0t}$ can be complex if required.

The characteristics of the reflected and the transmitted waves are obtained from the interface continuity conditions that the tangential components of $\mathbf{E}$ and $\mathbf{H}$ must be continuous across the interface, i.e. the sum of the tangential components of $\mathbf{E}_i$ and $\mathbf{E}_r$ must equal that of $\mathbf{E}_t$ on the interface ($z = 0$). Similar condition holds for $\mathbf{H}$. These conditions must hold for all instants of time and at all points on the interface ($z = 0$),

$$\therefore \quad \omega_i = \omega_r = \omega_t \tag{17.91}$$

i.e. all the three waves must be of same frequency.

Also, since these conditions hold at all the points on the interface,

$$\therefore \quad \frac{\mathbf{n}_i \cdot \mathbf{r}_i}{u_1} = \frac{\mathbf{n}_r \cdot \mathbf{r}_i}{u_1} = \frac{\mathbf{n}_t \cdot \mathbf{r}_i}{u_2} \tag{17.92}$$

where $\mathbf{r}_i$ is any point on the interface.

From the first two terms of the above equation

$$(\mathbf{n}_i - \mathbf{n}_r) \cdot \mathbf{r}_i = 0 \tag{17.93}$$

Since $\mathbf{r}_i$ lies on the interface, the vector $(\mathbf{n}_i - \mathbf{n}_r)$ must be normal to the interface plane $z = 0$, i.e. referring to the Figure 17.6,

$$\theta_i = \theta_r \tag{17.94}$$

which means that the angle of reflection equals the angle of incidence. Since $(\mathbf{n}_i - \mathbf{n}_r)$ is parallel to $\mathbf{n}$ (the normal to the interface), the three vectors $\mathbf{n}_i$, $\mathbf{n}_r$ and $\mathbf{n}$ are coplanar. These are the laws of reflection of the waves.

The plane containing these three vectors is called the 'plane of incidence'.

Going back to Eq. (17.92), we get

$$\left(\frac{\mathbf{n}_i}{u_1} - \frac{\mathbf{n}_t}{u_2}\right) \cdot \mathbf{r}_i = 0 \tag{17.95}$$

$\therefore$ The vector in the brackets must be normal to the interface plane $z = 0$, and hence $\mathbf{n}_i$, $\mathbf{n}_t$ and $\mathbf{n}$ are coplanar, and hence all the four normal vectors $\mathbf{n}_i$, $\mathbf{n}_t$, $\mathbf{n}_r$ and $\mathbf{n}$ must lie in the plane of incidence. Furthermore, the tangential components of $(\mathbf{n}_i/u_1)$ and $(\mathbf{n}_t/u_2)$ must be equal, i.e.

$$\frac{\sin \theta_i}{u_1} = \frac{\sin \theta_t}{u_2} \tag{17.96}$$

or since the wave number $\beta = \omega/u$, the above equality can be expressed as

$$\beta_1(\sin \theta_i) = \beta_2(\sin \theta_t)$$

or

$$\frac{\sin \theta_t}{\sin \theta_i} = \frac{\beta_1}{\beta_2} = \frac{n_1}{n_2} = \frac{\sqrt{\mu_{r1}\varepsilon_{r1}}}{\sqrt{\mu_{r2}\varepsilon_{r2}}} \tag{17.97}$$

where n is the index of refraction. This is the 'Snell's law of refraction'. Thus the quantity '$\beta \sin \theta$' is conserved across the interface. The above laws are general, and apply to any two media. They hold true for total reflection as well.

17.9 NORMAL INCIDENCE OF PLANE WAVES AT A PERFECTLY CONDUCTING BOUNDARY

We consider a plane-polarized, plane wave, incident normally on a perfectly conducting sheet. The direction of propagation is along the z-axis. The single frequency, uniform wave hits the plane, perfect conductor at $z = 0$. Now, there must be a reflected wave in addition to the incident wave, the reason being that a single travelling wave cannot satisfy the boundary condition on the perfect conducting surface on which the **E** field must be zero at all instants of time. (Also no energy can pass through the perfect conductor.) And hence all the energy from the incident wave must be returned by the reflected wave. Hence the incident and the reflected waves would be of equal amplitude, and together they form a standing wave pattern, whose properties we shall study now.

In our standard mathematical notations, the incident waves of **E** and **H** representing sinusoidal plane waves, moving towards $z \to +\infty$ are given by

$$E_{x+} = E_+ = \hat{E}_+ \cos \omega\left(t - \frac{z}{c}\right), \qquad B_{y+} = B_+ = \hat{B}_+ \cos \omega\left(t - \frac{z}{c}\right)$$

where

$$\hat{B}_+ = \frac{\hat{E}_+}{c} \tag{17.98}$$

Expressed as phasors, these are

$$\{E_{x+}\} = \hat{E}_+ \exp\left(\frac{-j\omega z}{c}\right) \qquad \text{and} \qquad \{B_{y+}\} = \hat{B}_+ \exp\left(\frac{-j\omega z}{c}\right)$$

As before, we write β (= the wave number) for (ω/c), and omit { } for the phasors on the understanding that all quantities are phasors. Thus

$$E_{x+} = E_+ = \hat{E}_+ \exp(-j\beta z), \; B_{y+} = B_+ = \hat{B}_+ \exp(-j\beta z), \; \left(\hat{B}_+ = \frac{\hat{E}_+}{c}\right) \tag{17.99}$$

This wave approaches the perfectly conducting boundary occupying the plane $z = 0$ from the negative side (Figure 17.7). **E** and **B** are both tangential to the interface. At the surface of the perfect conductor, the electric force is zero at all time, and hence there must be a reflected wave E_- such that $E_+ + E_- = 0$ at $z = 0$.

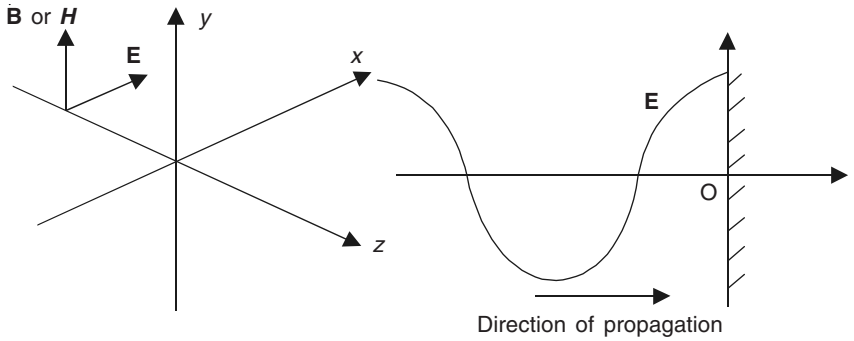


Figure 17.7 Incident E wave meeting $z = 0$ perfectly conducting surface, normally.

$$\therefore \quad E_- = -\hat{E}_+ \exp(j\beta z) \tag{17.100a}$$

and the associated magnetic vector B_- will have the same direction as B_+ (to give the propagation towards $z \to -\infty$) so that

$$B_- = \hat{B}_+ \exp(j\beta z) \tag{17.100b}$$

The total field is therefore

$$E_+ + E_- = \hat{E}_+ [\exp(-j\beta z) - \exp(j\beta z)] = -2j\hat{E}_+ \sin \beta z \tag{17.101a}$$

$$B_+ + B_- = \hat{B}_+ [\exp(-j\beta z) + \exp(j\beta z)] = 2j\hat{B}_+ \cos \beta z \tag{17.101b}$$

i.e. the travelling wave pattern has thus been replaced by a stationary pattern called a *standing wave*. It is such that $z = 0$ is always a zero or a *node* in the E-pattern, and a maximum or an *anti-node* in the B-pattern. Note that $E_+ + E_-$ and $B_+ + B_-$ are in space-quadrature; also that the power flow in the z-direction is zero (Figure 17.8).

If we need the **H** wave, this can be obtained either from the **B** wave or the **E** wave using the characteristic impedance of the medium, i.e.

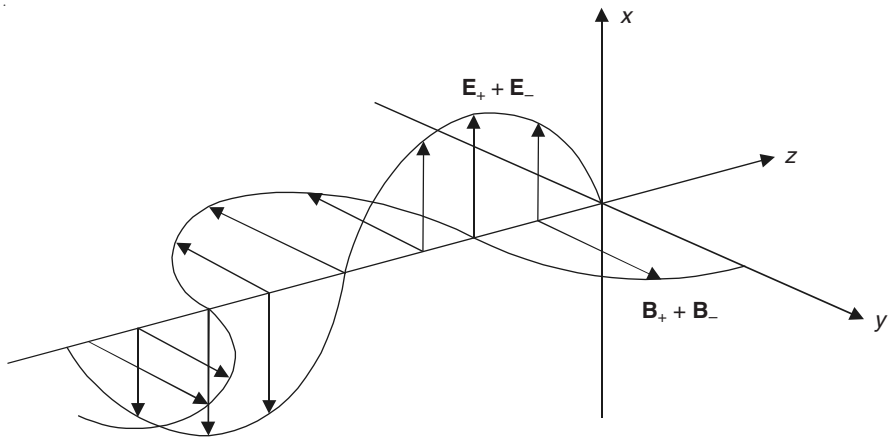


Figure 17.8 Standing wave patterns for **E** and **B**.

$$\frac{E_{x+}}{H_{y+}} = -\frac{E_{y+}}{H_{x+}} = Z_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}}$$

and

$$\frac{E_{x-}}{H_{y-}} = -\frac{E_{y-}}{H_{x-}} = -Z_0 = -\sqrt{\frac{\mu_0}{\varepsilon_0}}$$

$$\therefore \quad H_y = \left(\frac{E_{x+}}{Z_0}\right)\exp\left[j(\omega t - \beta z)\right] - \left(\frac{E_{x-}}{Z_0}\right)\exp\left[j(\omega t + \beta z)\right]$$

$$= \left(\frac{E_{x+}}{Z_0}\right)\left[\exp(-j\beta z) + \exp(j\beta z)\right]\exp(j\omega t)$$

$$= \left(\frac{2E_{x+}}{Z_0}\right)(\cos \beta z)\exp(j\omega t) \tag{17.102}$$

This is a standing wave pattern, similar to the **B** wave. Thus **E** and **H** patterns are mutually perpendicular in space (i.e. in space quadrature), and related in magnitude by Z_0; and they are in time-quadrature.

For the standing wave pattern of the **E** vector, **E** = 0 on the conductor surface ($z = 0$) at all t, and also for the values of $\beta z = -n\pi$, where n is an integer.

$\therefore$ Successive nodes (or zeroes) are located at

$$z = -\frac{n\pi}{\beta} = -\frac{n\lambda}{2} \tag{17.103}$$

where λ is the wavelength of the pattern.

H has maxima at the conductor surface and at all points where **E** = 0, i.e. $z = -n\lambda/2$.

Zeroes of **H** and maxima of **E** are at points where

$$\beta z = \frac{-(2n+1)\pi}{2} \qquad \text{or} \qquad z = \frac{-(2n+1)\pi}{2\beta} = \frac{-(2n+1)\lambda}{4} \tag{17.104}$$

$\therefore$ At an instant of time, occurring twice in each cycle, all the energy is in the magnetic field, and then 90° later all the energy is in the electric field.

The average value of the Poynting vector (**E** × **H**) is zero at each cross-sectional plane.

$\therefore$ On the average, as much energy is carried by the reflected wave as is brought by the incident wave.

The wave produces a surface current sheet on the interface—which is a perfect conductor.

$$J_{Sx} = H_y = \frac{2E_{x+}}{Z_0} \tag{17.105}$$

The current is in the +ve x-direction and the unit of its magnitude [given above in Eq. (17.105)] is amps/metre width in the y-direction.

17.10 REFLECTION FROM A DIELECTRIC (WITH NORMAL INCIDENCE)

We next consider the discontinuity due to two dielectrics, both being loss-less, of characteristic properties $\mu_1 (= \mu_0\mu_{r1})$, $\varepsilon_1 (= \varepsilon_0\varepsilon_{r1})$ and $\mu_2 (= \mu_0\mu_{r2})$, $\varepsilon_2 (= \varepsilon_0\varepsilon_{r2})$ respectively. A wave is incident

on the interface $z = 0$ normally from the region 1, the direction of propagation being the $+z$-direction (Figure 17.9). The resultant pattern would be:

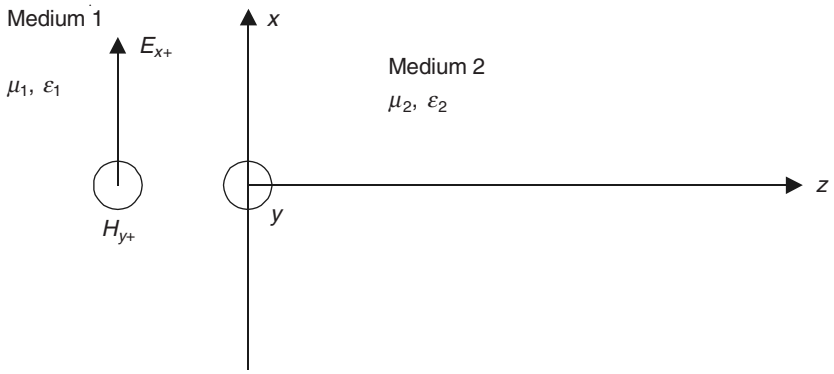


Figure 17.9 A plane wave incident on a dielectric interface.

in the medium 1: incident and reflected waves;

and

in the medium 2: transmitted wave only.

In the medium 1, the incident waves are

$$\mathbf{E}_{1+}(z) = \mathbf{i}_x E_{10+} \exp\,[\,j\,(\omega t - \beta_1 z)] \tag{17.106a}$$

$$\mathbf{H}_{1+}(z) = \mathbf{i}_y H_{10+} \exp\,[\,j\,(\omega t - \beta_1 z)]$$

$$= \mathbf{i}_y \left(\frac{E_{10+}}{Z_1}\right) \exp\,[\,j\,(\omega t - \beta_1 z)] \tag{17.106b}$$

where $\beta_1 = \omega\sqrt{\mu_1 \varepsilon_1}$, $Z_1 = \sqrt{\mu_1/\varepsilon_1}$, and both the waves are travelling in the $+z$-direction.

The reflected waves, travelling in the $-z$-direction, in the medium 1 are

$$\mathbf{E}_{1-}(z) = \mathbf{i}_x E_{10-} \exp\,[\,j\,(\omega t + \beta_1 z)] \tag{17.107a}$$

$$\mathbf{H}_{1-}(z) = \mathbf{i}_y H_{10-} \exp\,[\,j\,(\omega t + \beta_1 z)]$$

$$= \mathbf{i}_y \left(\frac{E_{10-}}{Z_1}\right) \exp\,[\,j\,(\omega t + \beta_1 z)] \tag{17.107b}$$

Also the transmitted waves would travel in the $+z$-direction, in the medium 2:

$$\mathbf{E}_{2+}(z) = \mathbf{i}_x E_{20+} \exp\,[\,j\,(\omega t - \beta_2 z)] \tag{17.108a}$$

$$\mathbf{H}_{2+}(z) = \mathbf{i}_y H_{20+} \exp\,[\,j\,(\omega t - \beta_2 z)]$$

$$= \mathbf{i}_y \left(\frac{E_{20+}}{Z_2}\right) \exp\,[\,j\,(\omega t - \beta_2 z)] \tag{17.108b}$$

where

$$\beta_2 = \omega\sqrt{\mu_2 \varepsilon_2}, \quad Z_2 = \sqrt{\frac{\mu_2}{\varepsilon_2}}$$

The unknowns in these equations are E_{10-} and E_{20+}, and the known specified quantity is E_{10+}.

The boundary conditions on the interface are:
On the plane $z = 0$,
(1) tangential **E** is continuous, i.e.

$$\mathbf{E}_{t1} = \mathbf{E}_{t2} \rightarrow E_{x1} = E_{x2}$$

and

(2) tangential **H** is continuous (as there is no surface current sheet), i.e.

$$\mathbf{H}_{t1} = \mathbf{H}_{t2} \rightarrow H_{y1} = H_{y2}$$

From the boundary conditions, we get

$$E_{20+} = \left(\frac{2Z_2}{Z_2 + Z_1}\right) E_{10+} \quad \text{and} \quad E_{10-} = \left(\frac{Z_2 - Z_1}{Z_2 + Z_1}\right) E_{10+} \tag{17.109}$$

The solution of the problem is:
In the medium 1:

$$\mathbf{E}_1 = \mathbf{i}_x E_{10+} \left[\exp(-j\beta_1 z) + \left(\frac{Z_2 - Z_1}{Z_2 + Z_1}\right) \exp(j\beta_1 z)\right] \exp(j\omega t) \tag{17.110a}$$

$$\mathbf{H}_1 = \mathbf{i}_x \left(\frac{E_{10+}}{Z_1}\right) \left[\exp(-j\beta_1 z) - \left(\frac{Z_2 - Z_1}{Z_2 + Z_1}\right) \exp(j\beta_1 z)\right] \exp(j\omega t) \tag{17.110b}$$

$$\mathbf{E}_2 = \mathbf{i}_x E_{10+} \left[\left(\frac{2Z_2}{Z_2 + Z_1}\right) \exp(-j\beta_2 z)\right] \exp(j\omega t) \tag{17.110c}$$

$$\mathbf{H}_2 = \mathbf{i}_x E_{10+} \left[\left(\frac{2}{Z_2 + Z_1}\right) \exp(-j\beta_2 z)\right] \exp(j\omega t) \tag{17.110d}$$

It should be noted that, in general, any discontinuity at a boundary will cause a reflected wave. In the present problem, the fraction of the incident wave, that is reflected, is

$$\frac{\mathbf{E}_{1-}}{\mathbf{E}_{1+}} = \frac{Z_2 - Z_1}{Z_2 + Z_1}$$

$$= \rho_E \text{ (the reflection coefficient)} \tag{17.111}$$

The above equation indicates that at the interface between two infinite dielectrics, the reflections can be eliminated only if $Z_2 = Z_1$. The fraction of the incident wave that is transmitted into the medium 2 is

$$\frac{\mathbf{E}_{2+}}{\mathbf{E}_{1+}} = \frac{2Z_2}{Z_2 + Z_1}$$

$$= \tau_E \text{ (the transmission coefficient)} = 1 + \rho_E \tag{17.112}$$

Next we have a look at the energy associated with each wave.

$\therefore$ The Poynting vector associated with the incident radiation is

$$\mathbf{S}_{1+av} = \mathbf{i}_z \left|\mathbf{E}_{1+}\right| \left|\mathbf{H}_{1+}\right| \cos\theta = \mathbf{i}_z \frac{\left|\mathbf{E}_{1+}\right|^2}{Z_1} = \mathbf{i}_z \frac{\left|E_{10+}\right|^2}{Z_1} \tag{17.113}$$

θ being the time-phase angle between **E** and **H** vectors of the incident waves. For the present case, E_x and H_y are in time-phase, and hence $\theta = 0$.

The Poynting vector associated with the reflected radiation is

$$\mathbf{S}_{1-\text{av}} = -\mathbf{i}_z \left[\frac{|E_{10+}|^2}{Z_1}\right]\left(\frac{Z_2 - Z_1}{Z_2 + Z_1}\right)^2 \tag{17.114}$$

and $\theta = 180°$ for this wave travelling in the $-z$-direction.

$\therefore$ The total $\mathbf{S}_{\text{av}}$ in the medium 1 = $\mathbf{S}_{1+\text{av}} + \mathbf{S}_{1-\text{av}}$

$$= \mathbf{i}_z \left[\frac{|E_{10+}|^2}{Z_1}\right]\left[1 - \left(\frac{Z_2 - Z_1}{Z_2 + Z_1}\right)^2\right]$$

$$= \mathbf{i}_z \left[\frac{|E_{10+}|^2}{Z_1}\right]\left[\frac{4Z_1Z_2}{(Z_2 + Z_1)^2}\right]$$

$$= \mathbf{i}_z |E_{10+}|^2 \left[\frac{4Z_2}{(Z_2 + Z_1)^2}\right] \tag{17.115}$$

In the medium 2:

The average value of the Poynting vector for the transmitted wave:

$$\mathbf{S}_{2+\text{av}} = \mathbf{i}_z \frac{|\mathbf{E}_{2+}|^2}{Z_2} = \mathbf{i}_z \frac{|E_{20+}|^2}{Z_2}$$

$$= \mathbf{i}_z \left[\frac{|E_{10+}|^2}{Z_1}\right]\left[\frac{4Z_1Z_2}{(Z_2 + Z_1)^2}\right]$$

$$= \mathbf{i}_z |E_{10+}|^2 \left[\frac{4Z_2}{(Z_2 + Z_1)^2}\right] \tag{17.116}$$

$\therefore$ There is no energy dissipation in the dielectrics at the interface.

17.11 THREE DIELECTRICS (NORMAL INCIDENCE)

The same method as discussed in Section 17.10 holds for any number of dielectrics. We consider three media as shown in Figure 17.10. A uniform plane wave: $\mathbf{E}_{\text{i}} = \mathbf{i}_x E_{1+} \exp [j(\omega t - \beta_1 z)]$ is incident from $z = -\infty$ on the first interface between the media 1 and 2. There would be transmitted and reflected waves at each interface for both **E** and **H** vectors, which would be mutually perpendicular, so that the resulting pattern would be given by the equations as stated below:

$$\mathbf{E}_{x1} = \mathbf{i}_x [E_{1+} \exp(-j\beta_1 z) + E_{1-} \exp(j\beta_1 z)] \exp(j\omega t) \tag{17.117a}$$

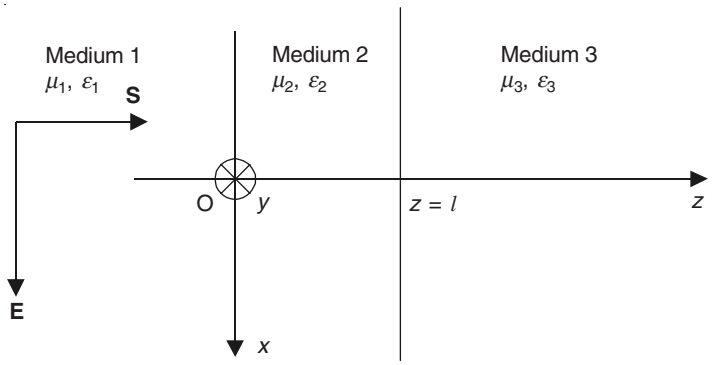


Figure 17.10 Plane wave in three dielectrics.

$$\mathbf{H}_{y1} = \mathbf{i}_y \left[\left(\frac{E_{1+}}{Z_1} \right) \exp\left(-j\beta_1 z\right) - \left(\frac{E_{1-}}{Z_1} \right) \exp\left(-j\beta_1 z\right) \right] \exp\,(j\omega t) \tag{17.117b}$$

$$\mathbf{E}_{x2} = \mathbf{i}_x \,[E_{2+} \exp\,(-j\beta_2 z) + \mathrm{E}_{2-} \exp\,(j\beta_2 z)] \exp\,(j\omega t) \tag{17.117c}$$

$$\mathbf{H}_{y2} = \mathbf{i}_y \left[\left(\frac{E_{2+}}{Z_2} \right) \exp\left(-j\beta_2 z\right) - \left(\frac{E_{2-}}{Z_2} \right) \exp\left(-j\beta_2 z\right) \right] \exp\,(j\omega t) \tag{17.117d}$$

$$\mathbf{E}_{x3} = \mathbf{i}_x E_{3+} \exp\,(-j\beta_3 z) \exp\,(j\omega t) \tag{17.117e}$$

$$\mathbf{H}_{y3} = \mathbf{i}_y \left(\frac{E_{3+}}{Z_3} \right) \exp\,(-j\beta_3 z) \exp\,(j\omega t) \tag{17.117f}$$

The unknowns are E_{1-}, E_{2+}, E_{2-}, and E_{3+}.

The boundary conditions at the two interfaces give:

(1) On $z = 0$, $\mathbf{E}_{x1} = \mathbf{E}_{x2}$ and $\mathbf{H}_{y1} = \mathbf{H}_{y2}$, from which we get

$$E_{1+} + E_{1-} = E_{2+} + E_{2-} \qquad \text{and} \qquad E_{1+} - E_{1-} = \left(\frac{Z_1}{Z_2} \right)(E_{2+} - E_{2-})$$

where

$$Z_1 = \sqrt{\frac{\mu_1}{\varepsilon_1}} \qquad \text{and} \qquad Z_2 = \sqrt{\frac{\mu_2}{\varepsilon_2}}$$

(2) On $z = l$, $\mathbf{E}_{x2} = \mathbf{E}_{x3}$ and $\mathbf{H}_{y2} = \mathbf{H}_{y3}$, from which we get

$$E_{2+} \exp\,(-j\beta_2 l) + E_{2-} \exp\,(+j\beta_2 l) = E_{3+} \exp\,(-j\beta_3 l)$$

and

$$E_{2+} \exp\,(-j\beta_2 l) - E_{2-} \exp\,(+j\beta_2 l) = \left(\frac{Z_2}{Z_3} \right) E_{3+} \exp\,(-j\beta_3 l)$$

where

$$\beta_2 = \omega\sqrt{\mu_2 \varepsilon_2} = \frac{\omega}{\mathrm{u}_2} = \frac{2\pi}{\lambda_2}$$

$$\beta_3 = \omega\sqrt{\mu_3 \varepsilon_3} = \frac{\omega}{\mathrm{u}_3} = \frac{2\pi}{\lambda_3}$$

From these equations, the four unknowns come out to be:

$$E_{1-} = \left(\frac{E_{1+}}{K}\right)\left[\left(Z_2 - Z_1\right) + \left(Z_2 + Z_1\right)\left(\frac{Z_3 - Z_2}{Z_3 + Z_2}\right)\exp\left(-j2\beta_2 l\right)\right] \tag{17.118a}$$

$$E_{2+} = \frac{2E_{1+}\, Z_2}{K} \tag{17.118b}$$

$$E_{2-} = \left(\frac{2E_{1+}\, Z_2}{K}\right)\left(\frac{Z_3 - Z_2}{Z_3 + Z_2}\right)\exp\left(-j2\beta_2 l\right) \tag{17.118c}$$

$$E_{3+} = \left[\frac{4E_{1+}\, Z_2 Z_3}{K(Z_3 + Z_2)}\right]\exp\left[-j\left(\beta_2 - \beta_3\right)l\right] \tag{17.118d}$$

where

$$K = \left[\left(Z_2 + Z_1\right) + \left(Z_2 - Z_1\right)\left(\frac{Z_3 - Z_2}{Z_3 + Z_2}\right)\exp\left(-j2\beta_2 l\right)\right] \tag{17.118e}$$

Let us consider a special case of a slab of dielectric in a homogeneous medium of infinite extent, like a sheet of glass or polystyrene in free space (or air), then $\varepsilon_1 = \varepsilon_3$ and $Z_3 = Z_1$, and therefore, from Eqs. (17.118),

$$\frac{E_{1-}}{E_{1+}} = \frac{\left(Z_2^2 - Z_1^2\right)\left[1 - \exp\left(-j2\beta_2 l\right)\right]}{\left(Z_2 - Z_1\right)^2 - \left(Z_2 + Z_1\right)^2 \exp\left(-j2\beta_2 l\right)} \tag{17.119}$$

i.e. the reflections will be eliminated, if $Z_2 = Z_1$, or

$$\exp\,(-j2\beta_2 l) = 1 \qquad \text{or} \qquad 2\beta_2 l = 2n\pi$$

Since $\beta_2 = \omega/u_2 = (2\pi f)/(f\lambda_2)$, the required thickness of the slab = l, must be:

$$l = \frac{n\pi}{\beta_2} = \frac{n\lambda_2}{2}$$

where λ_2 is the wavelength in the medium 2, and n is an integer, i.e. all the reflections are eliminated when the thickness of the dielectric plate is one-half the wavelength or its multiple.

Going back to the general case of three dielectrics again, from the first equation of (17.118), the required condition for no reflections gets complicated by the presence of the phase angle term exp $(-j2\beta_2 l)$ in part of the numerator. So in this case a double constraint is required, i.e.

exp $(-j2\beta_2 l) = \pm 1$, which implies that:

$$2\beta_2 l = 2n\pi \qquad \text{or} \qquad 2\beta_2 l = (2n + 1)\pi$$

so that $l = (2n + 1)\lambda_2/4$ (in the latter case) (17.120)

i.e. l has to be a quarter wavelength [or $(3/4)\lambda_2$, $(5/4)\lambda_2$, ...) and the modified ratio then becomes

$$\frac{E_{1-}}{E_{1+}} = \frac{K_1}{K_2} = (Z_2^2 - Z_1 Z_3)\,(Z_2^2 + Z_1 Z_3) \tag{17.121}$$

where

$$K_1 = (Z_2 - Z_1) - (Z_2 + Z_1)\left(\frac{Z_3 - Z_2}{Z_3 + Z_2}\right)$$

$$K_2 = (Z_2 + Z_1) - (Z_2 - Z_1)\left(\frac{Z_3 - Z_2}{Z_3 + Z_2}\right)$$

so that for the condition of no reflection,

$$Z_2 = \sqrt{Z_1 Z_3} \tag{17.122}$$

Thus, for no reflections, the required conditions are:

(1) the thickness l of the second medium must be a quarter wavelength as measured in that medium.
(2) the characteristic impedance of the second medium must be equal to the geometric mean of the characteristic impedances of the other two media.

It should be noted that when the reflections are eliminated, the whole energy is transmitted, since we are considering loss-less media. A practical application of this technique is in the reduction of radar reflections. But the limitation of this method is that it eliminates, though completely, one frequency only. Another application of this technique is in optics, where the camera lenses are coated with one-quarter wavelength thick coatings to eliminate certain colour lights.

17.12 THREE REGION PROBLEM WITH A THIN CONDUCTING LAYER (WITH NORMAL INCIDENCE)

This configuration will be shown to be another way of eliminating reflections, but here we shall not transmit all the energy as in the previous arrangement, and instead absorb it. We consider the general problem as follows.

A plane-polarized and plane electromagnetic wave is incident normally on a perfectly conducting surface. A thin conducting sheet is placed at a distance l from this surface and parallel to it (Figure 17.11). This sheet is so thin that it can be assumed to be a plane, and its surface resistivity is ρ_S Ω/metre^2. The surface of the perfect conductor lies on the plane $z = l$ as shown in the coordinate system of the figure.

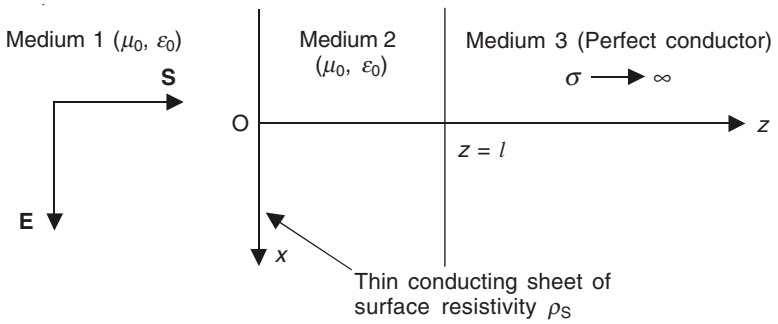


Figure 17.11 A plane wave normally incident on a thin conducting sheet, placed in front of a perfectly conducting plane.

The incident wave is = $\mathbf{E}_{i+} = \mathbf{i}_x E_{1+} \exp [j(\omega t - \beta z)]$
where $\beta = \omega/u = 2\pi/\lambda$, λ = wavelength, and $f = \omega/(2\pi)$ = frequency of repetition = $u/\lambda = c/\lambda$, since it is free space in this problem.

Across the conducting sheet (of finite surface resistivity), the wave will be partly reflected and partly transmitted.

∴ In both the regions 1 and 2, there will be both the reflected and the transmitted waves. But there will be no transmission through the perfectly conducting surface $z = l$ (of the region 3). The electric and the magnetic waves in the two regions would be:

$$\mathbf{E}_{1x} = \mathbf{i}_x [E_{1+} \exp(-j\beta z) + E_{1-} \exp(+j\beta z)] \exp(j\omega t) \tag{17.123a}$$

$$\mathbf{E}_{2x} = \mathbf{i}_x [E_{2+} \exp(-j\beta z) + E_{2-} \exp(+j\beta z)] \exp(j\omega t) \tag{17.123b}$$

$$\mathbf{H}_{1y} = \mathbf{i}_y [H_{1+} \exp(-j\beta z) + H_{1-} \exp(+j\beta z)] \exp(j\omega t) \tag{17.123c}$$

$$\mathbf{H}_{2y} = \mathbf{i}_y [H_{2+} \exp(-j\beta z) + H_{2-} \exp(+j\beta z)] \exp(j\omega t) \tag{17.123d}$$

The relationships between **E** and **H** are given by:

$$\left(\frac{E_{1+}}{H_{1+}}\right) = -\left(\frac{E_{1-}}{H_{1-}}\right) = \left(\frac{E_{2+}}{H_{2+}}\right) = -\left(\frac{E_{2-}}{H_{2-}}\right) = Z_0 = \sqrt{\mu_0/\varepsilon_0}$$

$$(\because \quad Z_1 = Z_2 = Z_0)$$

Hence,

$$\mathbf{H}_{1y} = \mathbf{i}_y \left[\left(\frac{E_{1+}}{Z_0}\right) \exp(-j\beta z) - \left(\frac{E_{1-}}{Z_0}\right) \exp(+j\beta z)\right] \exp(j\omega t) \tag{17.124a}$$

$$\mathbf{H}_{2y} = \mathbf{i}_y \left[\left(\frac{E_{2+}}{Z_0}\right) \exp(-j\beta z) - \left(\frac{E_{2-}}{Z_0}\right) \exp(+j\beta z)\right] \exp(j\omega t) \tag{17.124b}$$

The unknowns at this stage are: E_{1-}, E_{2+}, E_{2-}. To evaluate these, the requisite boundary conditions are:

(1) On the perfectly conducting surface $z = l$, $E_{2x} = 0$
(2) On the conducting sheet $z = 0$

(a) $E_{1x} = E_{2x}$ and (b) $\oint_C \mathbf{H} \cdot d\mathbf{l}$ = enclosed current (Figure 17.12), or $\nabla \times \mathbf{H} = \mathbf{J}$;

$$\therefore \quad (H_{1y} - H_{2y})\Delta w = \left(\frac{I_S}{w}\right)\Delta w$$

where (I_S/w) is the current in the conductor [in the x-direction perpendicular to the plane of the paper (Figure 17.12)] per unit width.

$$\therefore \quad (H_{1y} - H_{2y}) = \frac{E_{1x}}{\rho_S}$$

The three relevant equations obtained are:

$$E_{2x} = 0 = E_{2+} \exp(-j\beta l) + E_{2-} \exp(+j\beta l) \tag{17.125a}$$

$$E_{1+} + E_{1-} = E_{2+} + E_{2-} \tag{17.125b}$$

$$\left(\frac{E_{1+}}{Z_0} - \frac{E_{1-}}{Z_0}\right) - \left(\frac{E_{2+}}{Z_0} - \frac{E_{2-}}{Z_0}\right) = \frac{E_{1+} + E_{1-}}{\rho_S} \tag{17.125c}$$

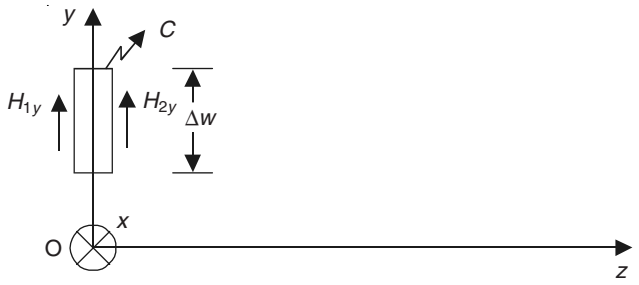


Figure 17.12 A closed contour C in the plane $z = 0$ (view of the y-z plane). Induced currents are in the x-direction into the plane of the paper.

From Eqs. (17.125a and b), we get

$$E_{2+} = \frac{E_{1+} + E_{1-}}{1 - \exp(-j2\beta l)}$$

and

$$E_{2-} = -\left[\frac{E_{1+} + E_{1-}}{1 - \exp(-j2\beta l)}\right] \exp(-j2\beta l)$$

Substituting these values for E_{2+} and E_{2-} in Eq. (17.125c),

$$E_{1+} - E_{1-} - \frac{E_{1+} + E_{1-}}{1 - \exp(-j2\beta l)} - \left[\frac{E_{1+} + E_{1-}}{1 - \exp(-j2\beta l)}\right] \exp(-j2\beta l) = \left(\frac{Z_0}{\rho_S}\right)(E_{1+} + E_{1-})$$

Solving this equation for E_{1-},

$$E_{1-} = \frac{E_{1+}\,[\exp(-j2\beta l)\,(Z_0 - 2\rho_S) - Z_a]}{2\rho_S + Z_0\,[1 - \exp(-j2\beta l)]}$$

To eliminate reflections, $E_{1-} = 0$, i.e. $[\exp(-j2\beta l)(Z_0 - 2\rho_S) - Z_0)] = 0$.

This is possible, only if $\exp(-j2\beta l)$ is real [i.e. $= \cos(2\beta l) - j \sin(2\beta l)$].

$\therefore$ $\sin(2\beta l) = 0$, i.e. $(2\beta l) = \pm \pi$; or $l = \pi/(2\beta) = (\pi/2)\{\lambda/(2\pi)\} = \lambda/4$.

$\therefore$ l must be $= \lambda/4$, or more generally $= (2n + 1)\lambda/4$, and $(Z_0 - 2\rho_S) - Z_0 = 0$, or $\rho_S = Z_0$. Thus all the incident power is absorbed by the conducting sheet only if the sheet has a surface resistivity equal to Z_0, and is placed one-quarter wavelength from a perfectly conducting plane.

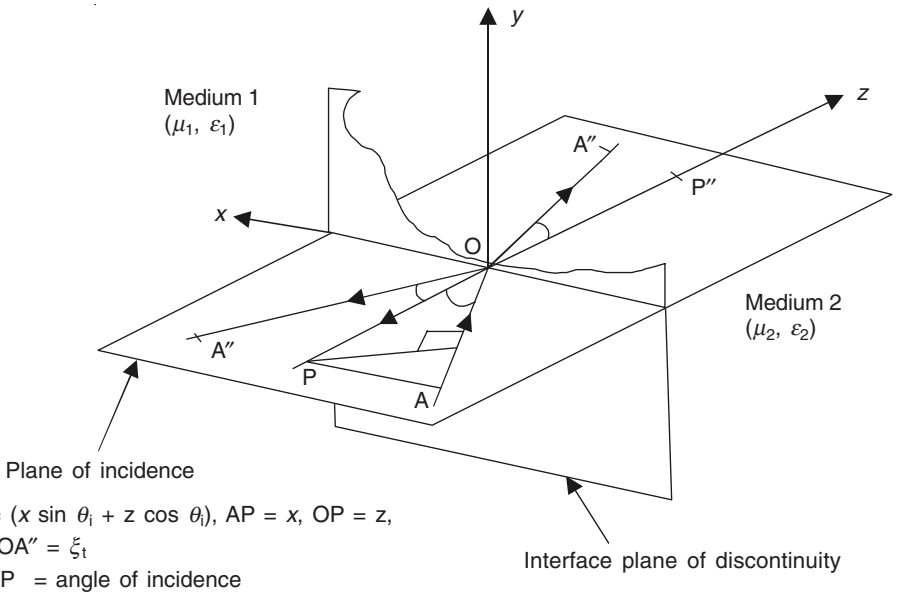
17.13 OBLIQUE INCIDENCE OF PLANE WAVES

So far we have considered a number of problems where the plane wave meets the interface of discontinuity normally, in which case the transmitted wave does not change its direction of propagation. We have seen in Section 17.8 that when the incident wave meets the interface of discontinuity, then the transmitted wave in the medium 2 changes its direction, i.e. the angle that the direction of propagation of the transmitted wave makes with the normal to the interface is different from the angle made by the direction of propagation of the incident wave with the same normal to the interface; thus, referring to Figure 17.6, $\theta_i \neq \theta_t$.

When the incident wave is normal to the interface, the mathematical representation of the orientation of the **E** and **H** vectors can always be simplified by a suitable choice of the coordinate system. But this is not so for the oblique incidence of the waves. For this purpose, we have to carefully define the plane of incidence (which we have already done before, but for convenience and clarity, we recapitulate here).

The plane of incidence is defined by a normal to the surface on which the wave impinges and a ray following the direction of propagation.

Following the notation of the direction vectors of the incident, the transmitted and the reflected waves as $\mathbf{n}_i$, $\mathbf{n}_t$, $\mathbf{n}_r$ and $\mathbf{n}$ (= normal to the interface), these were all proved to be coplanar, and the plane containing these vectors is being defined as the plane of incidence. We also remember that the **E** and **H** vectors in a plane electromagnetic wave are perpendicular to the direction of propagation and to each other. The **E** vector of the incident wave can then be oriented in any direction perpendicular to the vector $\mathbf{n}_i$. It is a normal practice to consider two orientations of the **E** vector, i.e. either **E** is parallel to the plane of incidence or **E** is normal to the plane of incidence. In fact any incident wave can be resolved into these two components. So we shall discuss the behaviour of the waves with these two orientations separately in different sections. However now the direction of propagation is no longer along a coordinate axis direction, and hence the equations for the three waves (i.e. the incident, the reflected and the transmitted) will now be somewhat more complicated, and the expressions for the exponents longer. Hence, if we choose an arbitrary point (x, z) on the direction of propagation of the wave, the equations for the waves can be written as (with respect to Figure 17.13, and following the results of Section 17.8)



OA = ξ_i = (x sin θ_i + z cos θ_i), AP = x, OP = z,
OA′ = ξ_t, OA″ = ξ_t
θ_i = ∠ AOP = angle of incidence
θ_r = ∠ A′OP = angle of reflection, $\theta_i = \theta_t$
θ_t = ∠ A″OP = angle of transmission
Arrows on the four rays indicate the unit vectors in the corresponding directions, i.e.
on OA—$\mathbf{n}_i$; on OA′—$\mathbf{n}_r$; on OA″—$\mathbf{n}_t$, and
on OP—**n** (the normal to the interface of discontinuity).

Figure 17.13 Coordinate system used for reflection and refraction of waves.

$$\mathbf{E}_i = \mathbf{E}_{0i} \exp\,[j\{\omega t - \beta_1(x \sin\,\theta_i + z \cos\,\theta_i)\}] \quad (17.126a)$$

$$\mathbf{E}_r = \mathbf{E}_{0r} \exp\,[j\{\omega t - \beta_1(-x \sin\,\theta_r + z \cos\,\theta_r)\}] \quad (17.126b)$$

$$\mathbf{E}_t = \mathbf{E}_{0t} \exp\,[j\{\omega t - \beta_2(x \sin\,\theta_t + z \cos\,\theta_t)\}] \quad (17.126c)$$

17.13.1 Oblique Incidence: Incident Wave Polarized with Its E Vector Normal to the Plane of Incidence

The **E** and **H** vectors of the incident wave are oriented as shown in Figure 17.14. Both the media are assumed to be isotropic and loss-less; and hence the **E** vectors of the reflected as well as the transmitted waves would also be normal to the plane of incidence as shown in the figure. The three waves can be written as

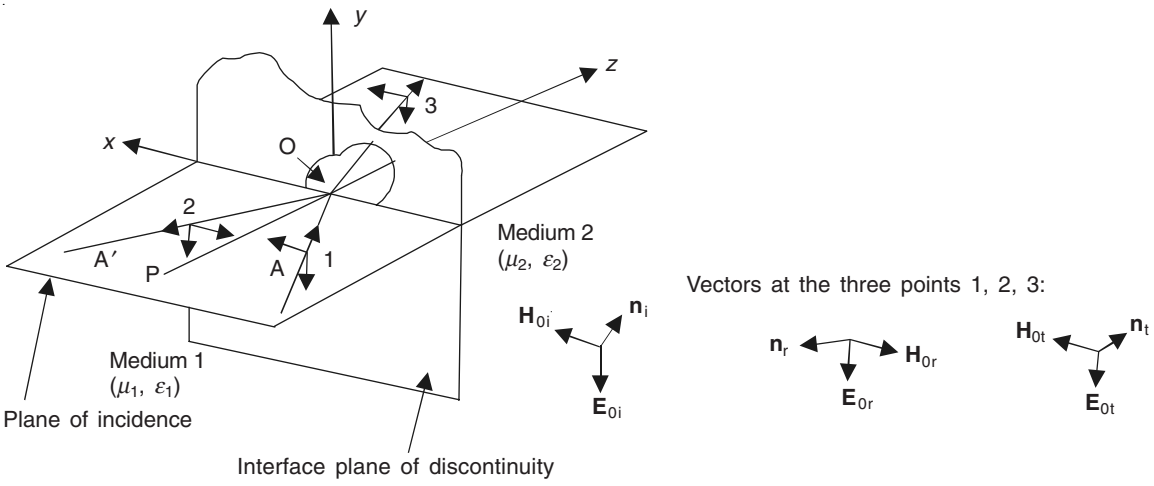


Figure 17.14 The incident, reflected and transmitted waves, when the incident wave has its **E** wave polarized normal to the plane of incidence.
(*Note:* All angles and lengths follow the same relationships as defined in Figure 17.13.)

$$\mathbf{E}_i = \mathbf{E}_{0i} \exp\,[j\{\omega t - \beta_1(x \sin\,\theta_i + z \cos\,\theta_i)\}] \quad (17.126a)$$

$$\mathbf{E}_r = \mathbf{E}_{0r} \exp\,[j\{\omega t - \beta_1(-x \sin\,\theta_r + z \cos\,\theta_r)\}] \quad (17.126b)$$

$$\mathbf{E}_t = \mathbf{E}_{0t} \exp\,[j\{\omega t - \beta_2(x \sin\,\theta_t + z \cos\,\theta_t)\}] \quad (17.126c)$$

where in terms of the lengths of the rays ξ_i, ξ_r, ξ_t, these are $\mathbf{E}_{0i} \exp\,[j\,(\omega t - \beta_1\xi_i)]$, $\mathbf{E}_{0r} \exp\,[j\,(\omega t - \beta_1\xi_r)]$, and $\mathbf{E}_{0t} \exp\,[j\,(\omega t - \beta_2\xi_t)]$ respectively.

We now evaluate $\mathbf{E}_{0r}$ and $\mathbf{E}_{0t}$ in terms of the known $\mathbf{E}_{0i}$ from the continuity conditions on the interface $z = 0$. The other unknowns $\mathbf{H}_{0r}$ and $\mathbf{H}_{0t}$ can then be obtained from the characteristic impedance relations between **E** and **H** vectors.

The interface continuity conditions to be used for evaluating the unknowns are:

1. The tangential **E** is continuous across $z = 0$.
2. The tangential **H** is continuous across $z = 0$, since there is no surface current on the interface.

$\therefore$ These conditions have to be met for all x and at all t.

$$\therefore \quad E_{0i} + E_{0r} = E_{0t} \tag{17.127}$$

(since **E** vector is parallel to the interface plane, as can be seen from Figure 17.7)

and

$$H_{0i} \cos\ \theta_i + H_{0r} \cos\ \theta_r = H_{0t} \cos\ \theta_t \tag{17.128}$$

Now,

$$\frac{E_{0i}}{H_{0i}} = Z_1 = \sqrt{\frac{\mu_1}{\varepsilon_1}} = \frac{\mu_1}{\sqrt{\mu_1\varepsilon_1}} = \frac{\omega\mu_1}{\omega\sqrt{\mu_1\varepsilon_1}}$$

$$= \frac{\omega\mu_1}{\beta_1}$$

Similarly,

$$\frac{E_{0r}}{H_{0r}} = \frac{\omega\mu_1}{\beta_1} \quad \text{and} \quad \frac{E_{0t}}{H_{0t}} = \frac{\omega\mu_2}{\beta_2} \tag{17.129}$$

So Eq. (17.128) becomes

$$\left(\frac{\beta_1}{\mu_1}\right)(E_{0i} - E_{0r}) \cos\theta_i = \left(\frac{\beta_2}{\mu_2}\right) E_{0t} \cos\theta_t \tag{17.130}$$

Now,

$$\beta_1 = \omega\sqrt{\mu_1\varepsilon_1} = \omega\sqrt{\mu_0\mu_{r1}\varepsilon_0\varepsilon_{r1}} = \omega\sqrt{\mu_0\varepsilon_0}\ \sqrt{\mu_{r1}\varepsilon_{r1}} = \left(\frac{\omega}{c}\right) n_1$$

and similarly $\beta_2 = \left(\frac{\omega}{c}\right) n_2$, and hence Eq. (17.130) becomes

$$\left(\frac{n_1}{\mu_1}\right)(E_{0i} - E_{0r}) \cos\theta_i = \left(\frac{n_2}{\mu_2}\right) E_{0t} \cos\theta_t \tag{17.131}$$

Hence, combining Eqs. (17.131) and (17.127), we get

$$\left(\frac{E_{0r}}{E_{0i}}\right)_N = \frac{\left(\frac{n_1}{\mu_1}\right)\cos\theta_i - \left(\frac{n_2}{\mu_2}\right)\cos\theta_t}{\left(\frac{n_1}{\mu_1}\right)\cos\theta_i + \left(\frac{n_2}{\mu_2}\right)\cos\theta_t} \tag{17.132a}$$

$$\left(\frac{E_{0t}}{E_{0i}}\right)_N = \frac{2\left(\frac{n_1}{\mu_1}\right)\cos\theta_i}{\left(\frac{n_1}{\mu_1}\right)\cos\theta_i + \left(\frac{n_2}{\mu_2}\right)\cos\theta_t} \tag{17.132b}$$

where the subscript N indicates that E_{0i} is normal to the plane of incidence. These equations [(17.132a and b)] are two of the 'Fresnel's equations'. The other two will be deduced in the next section. The Fresnel's equations state the ratios of the amplitudes of the incident, reflected and the transmitted waves. They apply to any two media, and we shall find that they are valid even for total reflection.

17.13.2 Oblique Incidence: Incident Wave Polarized with Its E Vector Parallel to the Plane of Incidence

In this case the **E** vectors of all the three waves would lie in the plane of incidence as shown in Figure 17.15, and the **H** vectors would be normal to this plane and parallel to the interface plane of discontinuity between the dielectrics. We have again to use the similar interface continuity conditions, i.e. the continuity of tangential **E** and tangential **H** on the plane $z = 0$.

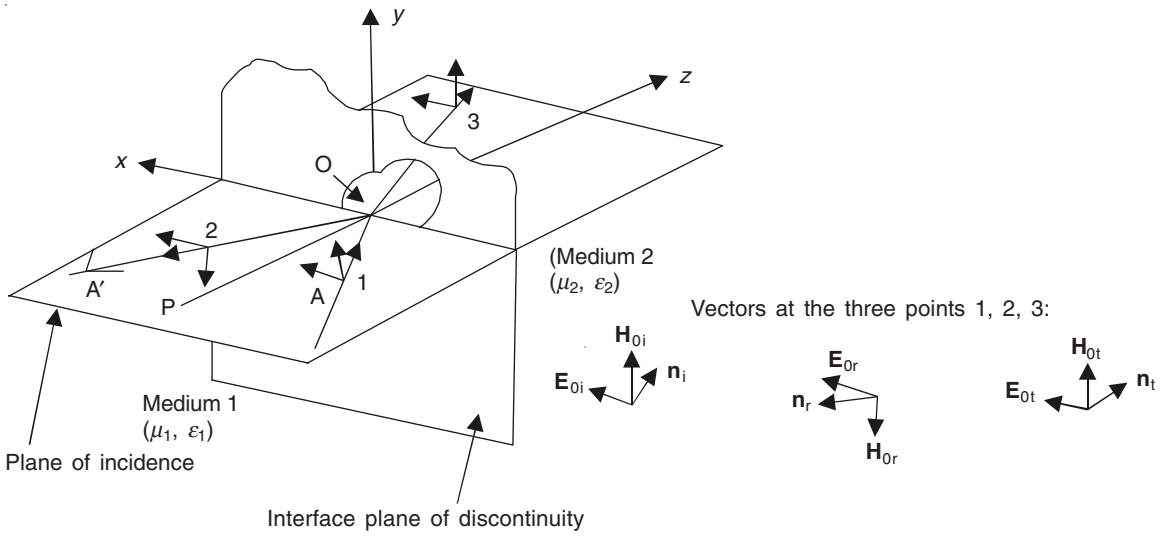


Figure 17.15 The incident, reflected, and transmitted waves, when the incident wave has its **E** wave polarized in the plane of incidence.
(*Note:* All the angles and the lengths follow the same relationships as in Figures 17.13 and 17.14. All subsequent similar figures in this chapter follow the same conventions as in these three diagrams, unless otherwise specifically stated.)

Applying the continuity conditions:

$$H_{0i} - H_{0r} = H_{0t} \qquad \text{or} \qquad \left(\frac{\beta_1}{\omega\mu_1}\right)(E_{0i} - E_{0r}) = \left(\frac{\beta_2}{\omega\mu_2}\right)E_{0t} \tag{17.133}$$

and

$$(E_{0i} + E_{0r}) \cos\theta_i = E_{0t} \cos\theta_t \tag{17.134}$$

Solving these two equations, as in the previous section, with the used substitutions for β in terms of the indices of refraction,

$$\left(\frac{E_{0r}}{E_{0i}}\right)_P = \frac{-\left(\frac{n_2}{\mu_2}\right)\cos\theta_i + \left(\frac{n_1}{\mu_1}\right)\cos\theta_t}{\left(\frac{n_2}{\mu_2}\right)\cos\theta_i + \left(\frac{n_1}{\mu_1}\right)\cos\theta_t} \tag{17.135a}$$

$$\left(\frac{E_{0t}}{E_{0i}}\right)_P = \frac{2\left(\frac{n_1}{\mu_1}\right)\cos\theta_i}{\left(\frac{n_2}{\mu_2}\right)\cos\theta_i + \left(\frac{n_1}{\mu_1}\right)\cos\theta_t} \tag{17.135b}$$

where the subscript P stands for E_{0i} being parallel to the plane of incidence. The four equations (17.132) and (17.135) are known as the 'Fresnel's equations'. When $\theta_i = \theta_t = 0$, we get the normal incidence of the incident wave, and the plane of incidence becomes indeterminate, and the pairs of equations (17.132) and (17.135) become identical; i.e.

$$\frac{E_{0r}}{E_{0i}} = \frac{\left(\frac{n_1}{\mu_1}\right) - \left(\frac{n_2}{\mu_2}\right)}{\left(\frac{n_1}{\mu_1}\right) + \left(\frac{n_2}{\mu_2}\right)} \tag{17.136a}$$

$$\frac{E_{0t}}{E_{0i}} = \frac{2\left(\frac{n_1}{\mu_1}\right)}{\left(\frac{n_1}{\mu_1}\right) + \left(\frac{n_2}{\mu_2}\right)} \tag{17.136b}$$

17.14 REFLECTION AND REFRACTION AT THE INTERFACE BETWEEN TWO NON-MAGNETIC DIELECTRICS (LOSS-LESS)

When we have an interface between two non-magnetic dielectrics, then $\mu_{r1} = \mu_{r2} = 1$, and then the equation for the Snell's law of refraction [Eq. (17.97)] becomes

$$\frac{\sin\theta_t}{\sin\theta_i} = \frac{n_1}{n_2} = \sqrt{\frac{\varepsilon_{r1}}{\varepsilon_{r2}}} \tag{17.137}$$

It should be noted that the index of refraction [Eq. (17.66)] would always be a number equal to or greater than unity. Also, the larger angle is in the medium with the lower index of refraction.

Considering a wave polarized with its **E** vector normal to the plane of incidence, the Fresnel's equations are:

$$\left(\frac{E_{0r}}{E_{0i}}\right)_N = \frac{\left(\frac{n_1}{n_2}\right)\cos\theta_i - \cos\theta_t}{\left(\frac{n_1}{n_2}\right)\cos\theta_i + \cos\theta_t} \tag{17.138a}$$

$$\left(\frac{E_{0t}}{E_{0i}}\right)_N = \frac{2\left(\frac{n_1}{n_2}\right)\cos\theta_i}{\left(\frac{n_1}{n_2}\right)\cos\theta_i + \cos\theta_t} \tag{17.138b}$$

It is seen that $(E_{0t}/E_{0i})_N$ is always real and positive. This means that at the interface, the transmitted wave is always in phase with the incident wave. On the other hand $(E_{0r}/E_{0i})_N$ can be either positive or negative depending on the value of (n_1/n_2). If, for example, $(n_1/n_2) > 1$, then $\theta_t > \theta_i$, and $\cos\theta_i > \cos\theta_t$. On the other hand, if $(n_1/n_2) < 1$, then $\theta_t < \theta_i$, and $\cos\theta_i < \cos\theta_t$. Hence the reflected wave is either in phase with the incident wave at the interface of discontinuity if $n_1 > n_2$, or is 180° out of phase if $n_1 < n_2$. Both these types of reflections are shown in Figure 17.16.

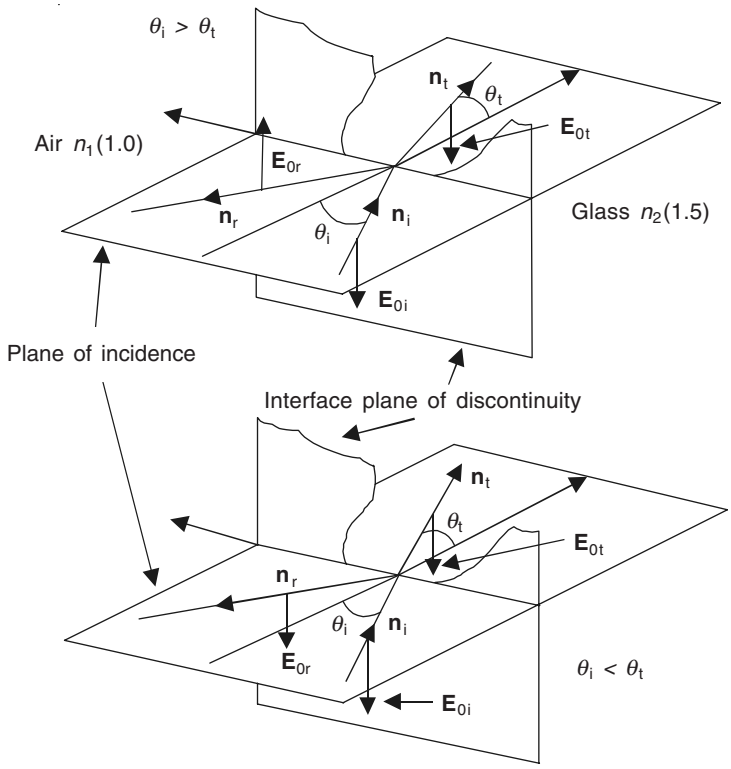


Figure 17.16 The relative phases, at the interface, of **E** in the transmitted and the reflected waves for $n_2 > n_1$ and $n_2 < n_1$ with E_{1oi} normal to the plane of incidence.

Next we consider the incident wave polarized with its **E** vector parallel to the plane of incidence. In this case, the Fresnel's equations become

$$\left(\frac{E_{0r}}{E_{0i}}\right)_P = \frac{-\cos\theta_i + \left(\frac{n_1}{n_2}\right)\cos\theta_t}{\cos\theta_i + \left(\frac{n_1}{n_2}\right)\cos\theta_t} \tag{17.139a}$$

$$\left(\frac{E_{0t}}{E_{0i}}\right)_P = \frac{2\left(\frac{n_1}{n_2}\right)\cos\theta_i}{\cos\theta_i + \left(\frac{n_1}{n_2}\right)\cos\theta_t} \tag{17.139b}$$

The second ratio (called the transmission coefficient) is again positive, which means that the incident $\mathbf{E}_i$ wave and the transmitted $\mathbf{E}_t$ wave are in phase (similar to the previous case, when **E** was normal to the plane of incidence).

But the reflection ratio (for $\mathbf{E}_{0r}$) can be either positive or negative, i.e. $\mathbf{E}_r$ and $\mathbf{E}_i$ can be either in phase or 180° out of phase (Figure 17.17).

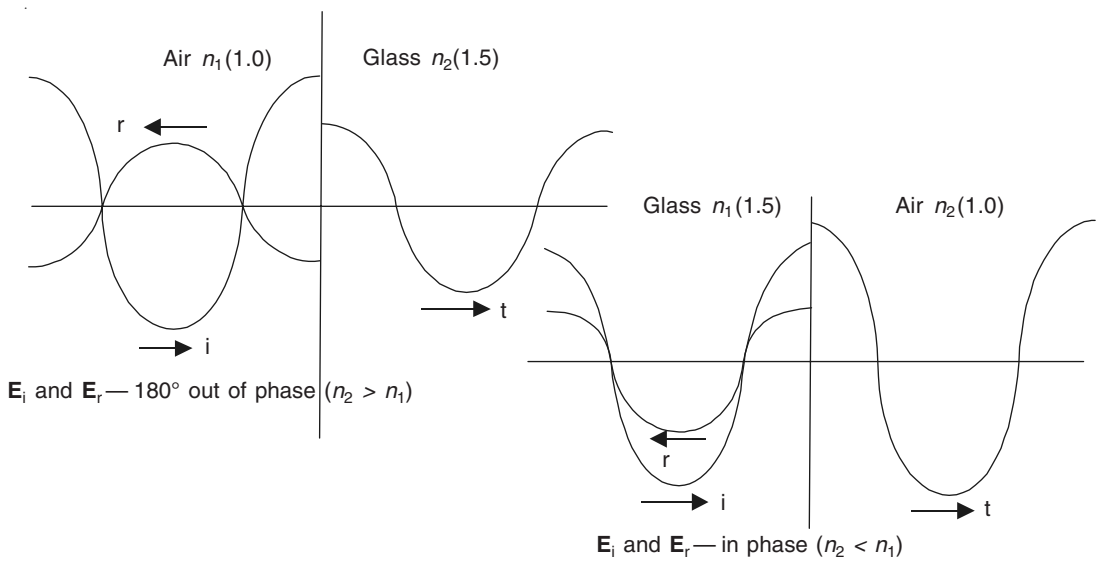


Figure 17.17 Graphs of $\mathbf{E}_i$, $\mathbf{E}_r$, $\mathbf{E}_t$ at normal incidence of the air-glass interface.

The $\mathbf{E}_{0r}$ will be in-phase with $\mathbf{E}_{0i}$ at the interface, if

$$\left(\frac{n_1}{n_2}\right)\cos\theta_t - \cos\theta_i > 0 \tag{17.140}$$

But $$\frac{n_1}{n_2} = \frac{\sin\theta_t}{\sin\theta_i} \quad \text{(Snell's law)}$$

$\therefore$ if $$\sin\theta_t\cos\theta_t - \sin\theta_i\cos\theta_i > 0$$

or $$\sin 2\theta_t - \sin 2\theta_i > 0$$

or if $$\sin(\theta_t - \theta_i)\cos(\theta_t + \theta_i) > 0 \tag{17.141}$$

this inequality is satisfied,

either if $\theta_t > \theta_i$ and $(\theta_t + \theta_i) < \pi/2$

or if $\theta_t < \theta_i$ and $(\theta_t + \theta_i) > \pi/2$ (17.142)

Thus the phase of the reflected wave does not depend only on the ratio of the refractive indices, i.e. (n_1/n_2); it also depends on both θ_t and θ_i. The ratio (E_{0r}/E_{0i}) can be either positive or negative, both for $n_2 > n_1$ and for $n_2 < n_1$, as seen in Eq. (17.142).

These ratios have been plotted, i.e. the reflected and the transmitted ratios for both the orientations (**E** vector normal to the plane of incidence, and then parallel to the plane of incidence) in Figure 17.18.

17.15 BREWSTER ANGLE AND TOTAL REFLECTION

From the discussion of Section 17.14, it is seen that in the polarization of the **E** vector parallel to the plane of incidence, there is an angle of incidence, called the 'Brewster angle', for which there is no reflected wave. This is seen from Eq. (17.141), when

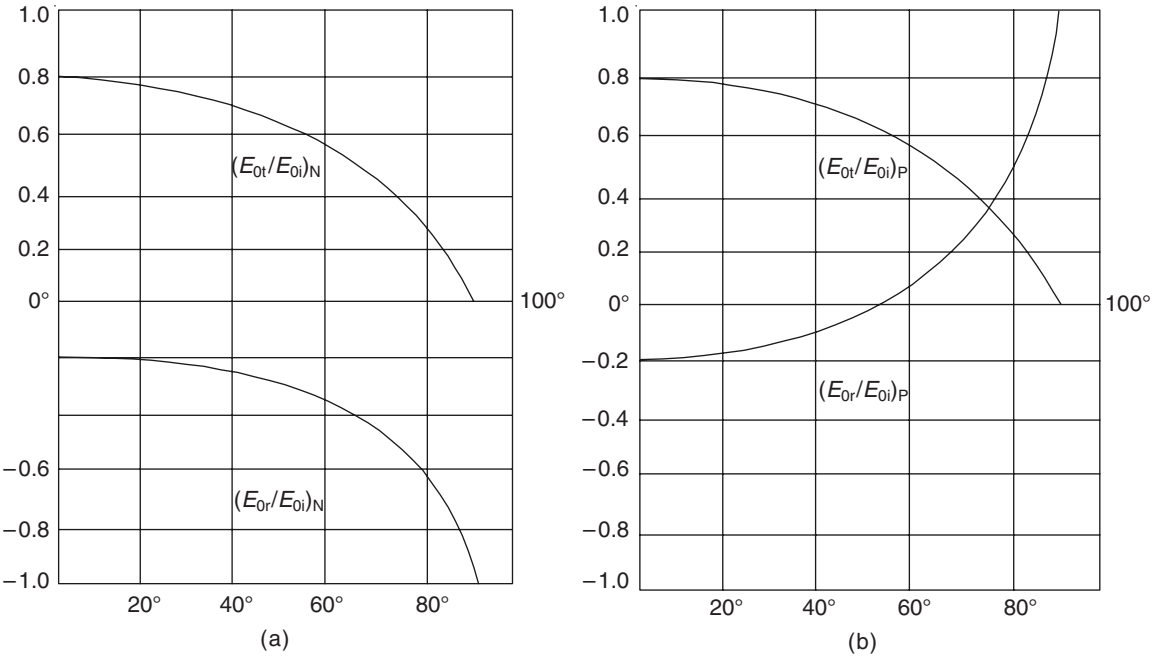


Figure 17.18 The ratios (E_{0t}/E_{0i}) and (E_{0r}/E_{0i}) as functions of the angle of incidence θ_i for $n_1/n_2 = 1.0/1.5$, for (a) **E** vector normal to the plane of incidence; and for (b) **E** vector parallel to the plane of incidence.

$$\sin(\theta_t - \theta_i)\cos(\theta_t + \theta_i) = 0$$

this gives the condition, which happens (i.e. no reflected wave) either when $\theta_i = \theta_t = 0$ or when $\theta_t + \theta_i = \pi/2$.

The first condition is wrong because this has been derived from the inequality $(n_1/n_2)\cos\theta_i - \cos\theta_t > 0$ being multiplied by $\sin\theta_i$ which is zero when $\theta_i = 0$.

∴ The condition for zero reflection is

$$\theta_i + \theta_t = \frac{\pi}{2} \tag{17.143}$$

This means that the conditions of continuity are satisfied by only two waves (for this situation), i.e. the incident wave and the transmitted wave as there is *no* reflected wave. This also implies that for this incident angle (= θ_{iB}), there would be a reflected wave only if the incident wave is polarized with its **E** vector normal to the plane of incidence. For this purpose, this angle of incidence (which is usually called the Brewster's angle) is also called the 'Polarizing angle', since an unpolarized wave incident on an interface at this angle is reflected as a polarized wave with its **E** vector normal to the plane of incidence. This angle is diagrammatically shown in Figure 17.19.

We shall now derive the expression for the Brewster's angle in terms of the media parameters. We start from the expression for the reflection ratio for the case of polarization with the **E** vector parallel to the plane of incidence, i.e. Eq. (17.139), which is

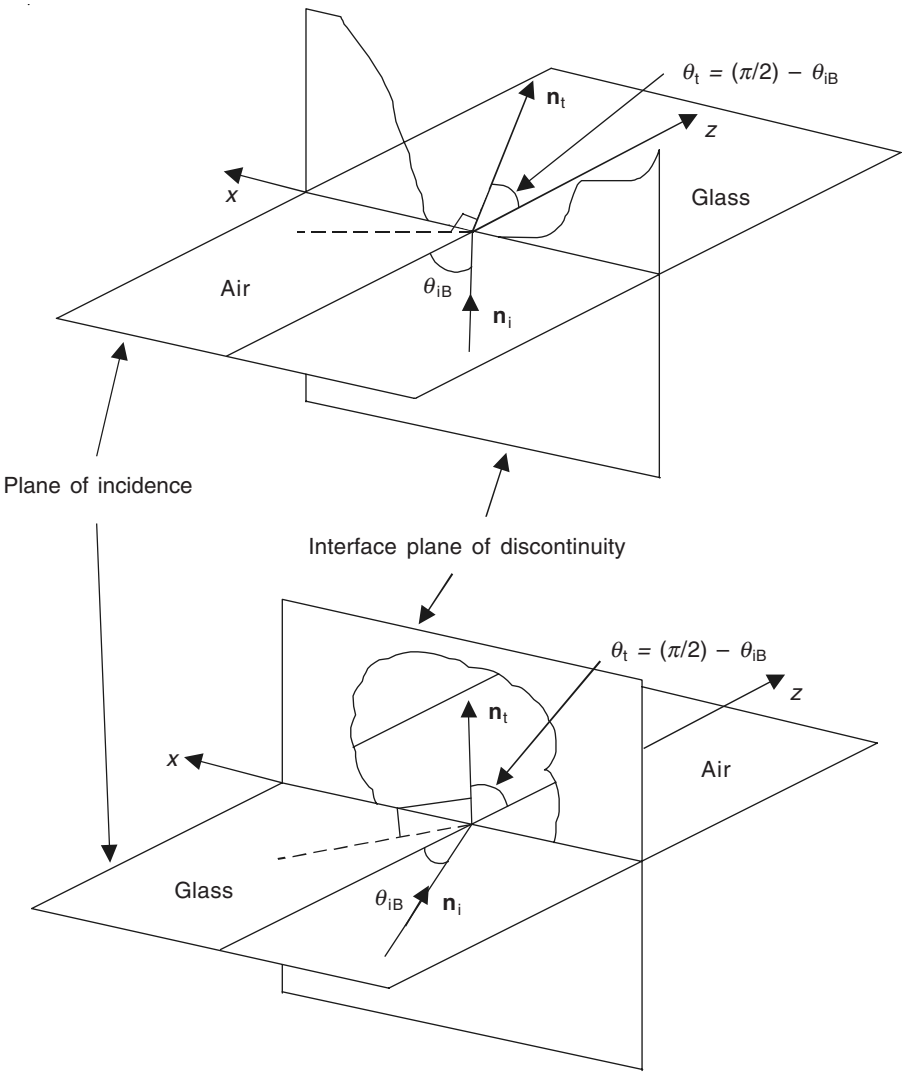


Figure 17.19 Brewster angle for air to glass and for glass to air.

$$\left(\frac{E_{0r}}{E_{0i}}\right)_P = \frac{-\cos\theta_i + \left(\frac{n_1}{n_2}\right)\cos\theta_t}{\cos\theta_i + \left(\frac{n_1}{n_2}\right)\cos\theta_t} \tag{17.139a}$$

and considering the more general expression of Section 17.13.2 (which imposes no constraint on the permeability of the media), i.e. Eq. (17.135) which states:

$$\left(\frac{E_{0r}}{E_{0i}}\right)_P = \frac{-\left(\frac{n_2}{\mu_2}\right)\cos\theta_i + \left(\frac{n_1}{\mu_1}\right)\cos\theta_t}{\left(\frac{n_2}{\mu_2}\right)\cos\theta_i + \left(\frac{n_1}{\mu_1}\right)\cos\theta_t} \tag{17.135a}$$

Since $n_1 = \sqrt{\mu_{r1}\varepsilon_{r1}}$ and $n_2 = \sqrt{\mu_{r2}\varepsilon_{r2}}$, substituting in the above equation, and multiplying the above equation by $\sqrt{\mu_0\varepsilon_0}$, we get

$$\left(\frac{E_{0r}}{E_{0i}}\right)_P = \frac{-\left(\frac{\sqrt{\mu_2\varepsilon_2}}{\mu_2}\right)\cos\theta_i + \left(\frac{\sqrt{\mu_1\varepsilon_1}}{\mu_1}\right)\cos\theta_t}{\left(\frac{\sqrt{\mu_2\varepsilon_2}}{\mu_2}\right)\cos\theta_i + \left(\frac{\sqrt{\mu_1\varepsilon_1}}{\mu_1}\right)\cos\theta_t}$$

The required condition for no reflection is

$$-\left(\frac{\sqrt{\mu_2\varepsilon_2}}{\mu_2}\right)\cos\theta_i + \left(\frac{\sqrt{\mu_1\varepsilon_1}}{\mu_1}\right)\cos\theta_t = 0$$

or

$$\left(\frac{\mu_2\varepsilon_2}{\mu_2^2}\right)\cos^2\theta_i = \left(\frac{\mu_1\varepsilon_1}{\mu_1^2}\right)\cos^2\theta_t = \left(\frac{\mu_1\varepsilon_1}{\mu_1^2}\right)\left(1 - \sin^2\theta_t\right) \tag{17.144}$$

From the Snell's law [Eq. (17.97)],

$$\sin\theta_t = \left(\frac{n_1}{n_2}\right)\sin\theta_i = \left(\sqrt{\frac{\mu_{r1}\varepsilon_{r1}}{\mu_{r2}\varepsilon_{r2}}}\right)\sin\theta_i$$

$$= \left(\sqrt{\frac{\mu_1\varepsilon_1}{\mu_2\varepsilon_2}}\right)\sin\theta_i$$

Simplifying Eq. (17.144), and substituting for θ_t,

$$(\mu_1\varepsilon_2)\cos^2\theta_i = (\mu_2\varepsilon_1)\left[1 - \frac{\mu_1\varepsilon_1}{\mu_2\varepsilon_2}\sin^2\theta_i\right]$$

or

$$\mu_1\varepsilon_2 - \mu_2\varepsilon_1 = \left(\mu_1\varepsilon_2 - \frac{\mu_1\varepsilon_1^2}{\varepsilon_2}\right)\sin^2\theta_i$$

$$\therefore \quad \sin^2\theta_i = \varepsilon_2\left[\frac{\mu_1\varepsilon_2 - \mu_2\varepsilon_1}{\mu_1(\varepsilon_2^2 - \varepsilon_1^2)}\right] \tag{17.145a}$$

and

$$\cos^2\theta_i = \varepsilon_1\left[\frac{\mu_2\varepsilon_2 - \mu_1\varepsilon_1}{\mu_1(\varepsilon_2^2 - \varepsilon_1^2)}\right] \tag{17.145b}$$

$$\therefore \quad \theta_i = \tan^{-1}\left[\frac{\varepsilon_2(\mu_1\varepsilon_2 - \mu_2\varepsilon_1)}{\varepsilon_1(\mu_2\varepsilon_2 - \mu_1\varepsilon_1)}\right]^{1/2} = \theta_{iB} \tag{17.145c}$$

This θ_{iB} is the Brewster's angle. For loss-less, non-magnetic dielectrics, $\mu_1 = \mu_2 = \mu_0$, this angle becomes

$$\theta_{iB} = \tan^{-1}\left(\sqrt{\frac{\varepsilon_2}{\varepsilon_1}}\right) = \tan^{-1}\left(\frac{n_2}{n_1}\right) \tag{17.146}$$

We shall now derive the Brewster's angle from the impedance considerations. This is the condition of 'perfect matching'. For this derivation, we refer to the equations of the Section 17.13.2 and Figure 17.15 of the same section. From that figure, the components of characteristic impedances from the incident and the transmitted waves are obtained as

$$Z_{iz} = \frac{E_{ix}}{H_{iy}} = \frac{E_{0i}\cos\theta_i}{H_{0i}} = Z_{0i}\cos\theta_i \tag{17.147a}$$

$$Z_{tz} = \frac{E_{tx}}{H_{ty}} = \frac{E_{0t}\cos\theta_t}{H_{0t}} = Z_{0t}\cos\theta_t \tag{17.147b}$$

For perfect matching,

$$Z_{iz} = Z_{tz} \tag{17.148}$$

or

$$Z_{0i}\cos\theta_i = Z_{0t}\cos\theta_t \tag{17.149}$$

We have

$$Z_{0i} = \sqrt{\frac{\mu_1}{\varepsilon_1}} \qquad \text{and} \qquad Z_{0t} = \sqrt{\frac{\mu_2}{\varepsilon_2}}$$

$\therefore$ From Eq. (17.149), we obtain

$$\left(\sqrt{\frac{\mu_1}{\varepsilon_1}}\right)\left(\sqrt{1-\sin^2\theta_i}\right) = \left(\sqrt{\frac{\mu_2}{\varepsilon_2}}\right)\left(\sqrt{1-\sin^2\theta_t}\right)$$

From the Snell's law [Eq. (17.97)],

$$\sin\theta_t = \left(\frac{n_1}{n_2}\right)\sin\theta_i = \left(\sqrt{\frac{\mu_{r1}\varepsilon_{r1}}{\mu_{r2}\varepsilon_{r2}}}\right)\sin\theta_i$$

$$= \left(\sqrt{\frac{\mu_1\varepsilon_1}{\mu_2\varepsilon_2}}\right)\sin\theta_i$$

$\therefore$ Substituting in the above, and squaring and cross-multiplying

$$(\mu_1\varepsilon_2)(1-\sin^2\theta_i) = \mu_2\varepsilon_1\left[1-\left(\frac{\mu_1\varepsilon_1}{\mu_2\varepsilon_2}\right)\sin^2\theta_i\right]$$

$$\therefore \quad \sin^2\theta_i = \frac{\varepsilon_2(\mu_1\varepsilon_2 - \mu_2\varepsilon_1)}{\mu_1(\varepsilon_2^2 - \varepsilon_1^2)}$$

which is the same as Eq. (17.145a).

Next we write the reflection coefficient ratio of Eq. (17.135) in terms of the characteristic impedances, i.e.

$$\left(\frac{E_{0r}}{E_{0i}}\right)_P = \frac{-\left(\frac{n_2}{\mu_2}\right)\cos\theta_i + \left(\frac{n_1}{\mu_1}\right)\cos\theta_t}{\left(\frac{n_2}{\mu_2}\right)\cos\theta_i + \left(\frac{n_1}{\mu_1}\right)\cos\theta_t} \tag{17.135a}$$

$$= \frac{-\left(\sqrt{\frac{\varepsilon_2}{\mu_2}}\right)\cos\theta_i + \left(\sqrt{\frac{\varepsilon_1}{\mu_1}}\right)\cos\theta_t}{\left(\sqrt{\frac{\varepsilon_2}{\mu_2}}\right)\cos\theta_i + \left(\sqrt{\frac{\varepsilon_1}{\mu_1}}\right)\cos\theta_t}$$

$$\therefore \quad \left(\frac{E_{0r}}{E_{0i}}\right)_P = \frac{Z_{02}\cos\theta_t - Z_{01}\cos\theta_i}{Z_{02}\cos\theta_t + Z_{01}\cos\theta_i} \tag{17.150}$$

This is a more useful form to derive the condition for perfect matching from impedance considerations.

The Brewster's angle corresponding to the parallel magnetic field (i.e. parallel to the plane of incidence), can also be similarly obtained, this being useful in radio wave reflections.

Before we go on to discuss the next topic of total reflection, we shall define the energy flow across the present interface. The average energy flow (or flux) per unit area in the incident wave is given by the average value of the Poynting vector as shown in Eq. (17.68). Hence,

$$\bar{\mathbf{S}}_{i\,av} = \frac{1}{2}\left(\sqrt{\frac{\varepsilon_1}{\mu_1}}\right)E_{0i}^2\mathbf{n}_i \tag{17.151a}$$

$$\bar{\mathbf{S}}_{r\,av} = \frac{1}{2}\left(\sqrt{\frac{\varepsilon_1}{\mu_1}}\right)E_{0r}^2\mathbf{n}_r \tag{17.151b}$$

$$\bar{\mathbf{S}}_{t\,av} = \frac{1}{2}\left(\sqrt{\frac{\varepsilon_2}{\mu_2}}\right)E_{0t}^2\mathbf{n}_t \tag{17.151c}$$

where $\mathbf{n}_i$, $\mathbf{n}_r$, and $\mathbf{n}_t$ are the unit vectors in the directions of incidence, reflection, and transmission respectively.

The *coefficients of energy reflection R* and *energy transmission T* are defined as the ratios of the average energy fluxes per unit time and per unit area at the interface, i.e.

$$R = \left|\frac{\bar{\mathbf{S}}_{r\,av}\cdot\mathbf{n}}{\bar{\mathbf{S}}_{i\,av}\cdot\mathbf{n}}\right| = \frac{E_{0r}^2}{E_{0i}^2} \tag{17.152a}$$

$$T = \left|\frac{\bar{\mathbf{S}}_{t\,av}\cdot\mathbf{n}}{\bar{\mathbf{S}}_{i\,av}\cdot\mathbf{n}}\right| = \left(\sqrt{\frac{\varepsilon_{r2}\mu_{r1}}{\varepsilon_{r1}\mu_{r2}}}\right)\frac{E_{0r}^2}{E_{0i}^2} \tag{17.152b}$$

where $\mathbf{n}$ is the unit vector normal to the interface.

If we consider the interface between two loss-less, non-magnetic dielectrics, then from the Fresnel's equations, we have

$$R_N = \left[\frac{\left(\frac{n_1}{n_2}\right)\cos\theta_i - \cos\theta_t}{\left(\frac{n_1}{n_2}\right)\cos\theta_i + \cos\theta_t} \right]^2 \tag{17.153a}$$

$$T_N = \frac{4\left(\frac{n_1}{n_2}\right)\cos\theta_i \cos\theta_t}{\left\{\left(\frac{n_1}{n_2}\right)\cos\theta_i + \cos\theta_t\right\}^2} \tag{17.153b}$$

$$R_P = \left[\frac{-\cos\theta_i + \left(\frac{n_1}{n_2}\right)\cos\theta_t}{\cos\theta_i + \left(\frac{n_1}{n_2}\right)\cos\theta_t} \right]^2 \tag{17.153c}$$

$$T_P = \frac{4\left(\frac{n_1}{n_2}\right)\cos\theta_i \cos\theta_t}{\left\{\cos\theta_i + \left(\frac{n_1}{n_2}\right)\cos\theta_t\right\}^2} \tag{17.153d}$$

As we are considering loss-less dielectrics, in both the cases, $R + T = 1$, as all the energy would be conserved.

At the Brewster's angle θ_{iB} [defined by Eq. (17.146)],

$$R_P = 0 \qquad \text{and} \qquad T_P = 1$$

17.15.1 Total Reflection

According to the Snell's law,

$$\frac{\sin\theta_t}{\sin\theta_i} = \sqrt{\frac{\mu_1\varepsilon_1}{\mu_2\varepsilon_2}} = \frac{n_1}{n_2} \tag{17.154}$$

If $(\mu_1\varepsilon_1) > (\mu_2\varepsilon_2)$, then $\sin\theta_t$ can be > 1, for a real value of $\sin\theta_i$ (which is apparently a mathematically absurd result). The critical angle of incidence for which $\sin\theta_t = 1$, and $\theta_t = \pi/2$, is

$$\sin\theta_{ic} = \frac{n_2}{n_1} \tag{17.155}$$

Experimentally it has been observed that when $\theta_i \geq \theta_{ic}$, the wave starting in the medium 1 and incident on the interface, is totally reflected back in the medium 1 (Figure 17.20). This phenomenon is called the 'total reflection', and does not depend on the orientation of the **E** vector of the incident wave. For the light, propagating in glass with an index of refraction of 1.6, the critical angle of incidence is 38.7°. It should be noted that the Snell's law, the laws of reflection and refraction, and the Fresnel's equations are all applicable to the total reflection (by disregarding the fact that $\sin\theta_t > 1$ for the time being, and writing as) i.e.

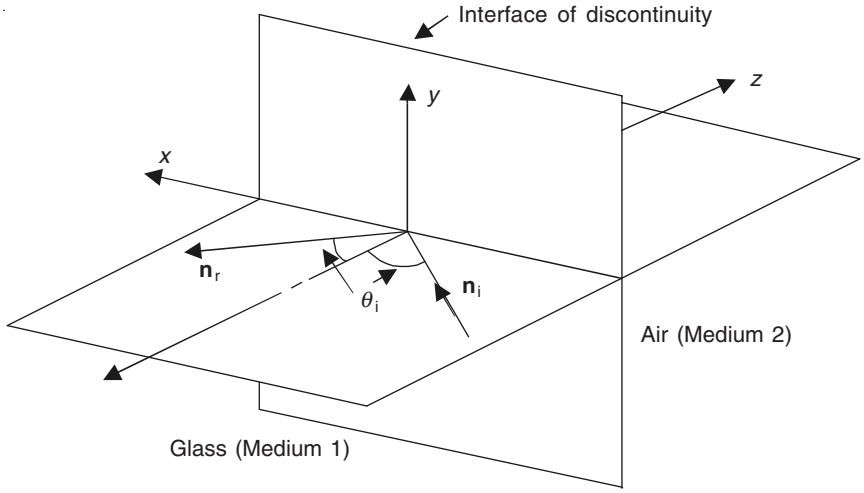


Figure 17.20 Total reflection.

$$\cos\theta_t = -\left(1-\sin^2\theta_t\right)^{1/2} = -\left[1-\left(\frac{n_1}{n_2}\right)^2\sin^2\theta_i\right]^{1/2}$$

$$= -j\left(\frac{n_1}{n_2}\right)\left[\sin^2\theta_i - \left(\frac{n_2}{n_1}\right)^2\right]^{1/2} \tag{17.156}$$

where $n_1 > n_2$.

Let us consider the incident wave

$$\mathbf{E}_{0i}\exp\left[j\{\omega t - \beta_1(x\sin\theta_i + z\cos\theta_i)\}\right]$$

And the transmitted wave

$$\mathbf{E}_{0t}\exp\left[j\{\omega t - \beta_2(x\sin\theta_t + z\cos\theta_t)\}\right]$$

So now the question is: what does it mean if $\sin\theta_t > 1$?

Let $\beta\sin\theta_t = \gamma$, then $\beta\cos\theta_t = \beta\sqrt{1-\sin^2\theta_t} = \pm j\alpha$

$\therefore$ The transmitted wave becomes

$$\mathbf{E}_{0t}\exp\left[j(\omega t - \gamma_2 x)\exp(\pm\alpha z)\right]$$

which means that the wave is travelling only in the x-direction (i.e. along the interface of the discontinuity), i.e. a surface wave, decreasing [or increasing for $\exp(+\alpha z)$] exponentially in the z-direction—normal to the plane $z = 0$. To understand this expression, we consider, in a more rigorous manner the expressions for the reflected and the transmitted waves when the **E** vector is normal to the plane of incidence, i.e.

$$\mathbf{E}_i = \mathbf{i}_y E_{0i}\exp\left[j\{\omega t - \beta_1(x\sin\theta_i + z\cos\theta_i)\}\right]$$

$$\mathbf{E}_r = \mathbf{i}_y E_{0r}\exp\left[j\{\omega t - \beta_1(-x\sin\theta_r + z\cos\theta_r)\}\right]$$

$$\mathbf{E}_t = \mathbf{i}_y E_{0t}\exp\left[j\{\omega t - \beta_2(x\sin\theta_t + z\cos\theta_t)\}\right]$$

From the Fresnel's equations (17.132),

$$\left(\frac{E_{0r}}{E_{0i}}\right)_N = \frac{\left(\frac{n_1}{\mu_1}\right)\cos\theta_i - \left(\frac{n_2}{\mu_2}\right)\cos\theta_t}{\left(\frac{n_1}{\mu_1}\right)\cos\theta_i + \left(\frac{n_2}{\mu_2}\right)\cos\theta_t} \tag{17.132a}$$

$$\left(\frac{E_{0t}}{E_{0i}}\right)_N = \frac{2\left(\frac{n_1}{\mu_1}\right)\cos\theta_i}{\left(\frac{n_1}{\mu_1}\right)\cos\theta_i + \left(\frac{n_2}{\mu_2}\right)\cos\theta_t} \tag{17.132b}$$

For non-magnetic, loss-less dielectrics,

$$\mu_1 = \mu_2 = \mu_0 \qquad \text{and} \qquad \cos\theta_t = -j\left(\frac{n_1}{n_2}\right)\left[\sin^2\theta_i - \left(\frac{n_2}{n_1}\right)^2\right]^{1/2}$$

Substituting in Eqs. (17.132) and simplifying,

$$\left(\frac{E_{0r}}{E_{0i}}\right)_N = \frac{\cos\theta_i + j\sqrt{\sin^2\theta_i - \left(\frac{n_2}{n_1}\right)^2}}{\cos\theta_i - j\sqrt{\sin^2\theta_i - \left(\frac{n_2}{n_1}\right)^2}}$$

$$\left(\frac{E_{0t}}{E_{0i}}\right)_N = \frac{2\cos\theta_i}{\cos\theta_i - j\sqrt{\sin^2\theta_i - \left(\frac{n_2}{n_1}\right)^2}}$$

Rationalizing, we get

$$\left(\frac{E_{0r}}{E_{0i}}\right)_N = \frac{\left[\cos\theta_i + j\sqrt{\sin^2\theta_i - \left(\frac{n_2}{n_1}\right)^2}\right]^2}{1 - \left(\frac{n_2}{n_1}\right)^2}$$

$$\left(\frac{E_{0t}}{E_{0i}}\right)_N = \frac{2\cos\theta_i\left[\cos\theta_i + j\sqrt{\sin^2\theta_i - \left(\frac{n_2}{n_1}\right)^2}\right]}{1 - \left(\frac{n_2}{n_1}\right)^2}$$

On further simplifying, these expressions finally reduce to:

$$\left(\frac{E_{0r}}{E_{0i}}\right)_N = \exp\left[j\ 2\ \tan^{-1}\left\{\frac{\sqrt{\sin^2\theta_i - \left(\frac{n_2}{n_1}\right)^2}}{\cos\theta_i}\right\}\right] = \exp(j\alpha) \tag{17.157a}$$

and

$$\left(\frac{E_{0t}}{E_{0i}}\right)_N = \frac{2\cos\theta_i}{1-\left(\frac{n_2}{n_1}\right)^2} \exp\left[j\ \tan^{-1}\left\{\frac{\sqrt{\sin^2\theta_i - \left(\frac{n_2}{n_1}\right)^2}}{\cos\theta_i}\right\}\right] \tag{17.157b}$$

It should be noted that the amplitude of the reflected wave is equal to the amplitude of the incident wave, so that the coefficient of reflection R is equal to unity. The energy is totally reflected, and the net flux of energy through the interface is zero.

From the above equations, it is also seen that the transmitted wave E_{0t} is not zero, though the net flux of the energy across the interface is zero. The medium 2 can be equated to a pure inductance, the average power to which is zero (the power flow being oscillatory, oscillating one way and then the other), but there is still a current (reactive type) through the inductance.

The transmitted wave is

$$\mathbf{E}_t = \mathbf{i}_y E_{0t} \exp\left[j\left\{\omega t - \left(\frac{n_1}{n_2}\right)\beta_2 x \sin\theta_i\right\} + \left(\frac{n_1}{n_2}\right)\beta_2 \left\{\sqrt{\sin^2\theta_i - \left(\frac{n_2}{n_1}\right)^2}\right\} z\right] \tag{17.158}$$

which travels unattenuated parallel to the interface. Its wavelength in the x-direction (parallel to the interface) is

$$\lambda_x = \frac{\lambda_0}{n_1 \sin\theta_i} = \frac{\lambda_1}{\sin\theta_i} \tag{17.159}$$

where λ_x is the distance along the x-axis between two neighbouring equiphase points in the incident wave.

This result that the wave travels unattenuated parallel to the interface is quite remarkable. The question as to whether the transmitted wave extends to the other side of the interface cannot be answered on the basis of the present discussion, which is based on the assumption that the incident wave extends to infinity. In reality, what happens is that the incident ray, instead of being reflected abruptly at the interface, penetrates into the medium 2, where it is bent back into the medium 1. The transmitted wave is damped exponentially in the direction normal to the interface (the z-direction), in such a way that its amplitude reduces by a factor e over a distance, which is

$$\delta_z = \frac{\lambda_2}{2\pi\left[\left(\frac{n_1}{n_2}\right)^2 \sin^2\theta_i - 1\right]^{1/2}} \tag{17.160}$$

It should be noted that the critical angle θ_c [= $\sin^{-1}(n_2/n_1)$] is somewhat larger than the Brewster angle θ_{iB} [= $\tan^{-1}(n_2/n_1)$]. For the glass-air interface, where the refractive index of the glass is 1.6, $\theta_c = 38.7°$ and $\theta_{iB} = 32°$.

17.15.2 Reflection and Refraction, and Fresnel's Equations in Total Reflection

It is impossible to satisfy the continuity conditions of the tangential **E** and **H** at the interface of discontinuity, with only the incident and the reflected waves. So we are forced to conclude that there would be some sort of transmitted wave, of a rather special nature, since it is not observable under ordinary conditions. It must, of course, satisfy the general wave equation for the non-magnetic dielectrics, i.e.

$$\frac{\partial^2 E_t}{\partial x^2} + \frac{\partial^2 E_t}{\partial z^2} = \mu_0 \varepsilon_2 \left(\frac{\partial^2 E_t}{\partial t^2} \right) \tag{17.161}$$

(In our coordinate system, for the configurations we have been considering, there is no variation in the y-direction, i.e. $\partial/\partial y = 0$). The incident wave has been defined as

$$\mathbf{E}_i = \mathbf{E}_{0i} \exp\,[j\{\omega t - \beta_1(x \sin\,\theta_i + z \cos\,\theta_i)\}] \tag{17.162}$$

The reflected wave can be written as

$$\mathbf{E}_r = \mathbf{E}_{0r} \exp\,[j(\omega t - \beta_{1x} x + \beta_{1z} z)] \tag{17.163}$$

where $\mathbf{E}_{0r}$, β_{1x}, β_{1z} are unknown constants. The reflected wave would have the same angular frequency ω as the incident wave. Let the transmitted wave be written as

$$\mathbf{E}_t = (\mathbf{i}_x E_{0tx} + \mathbf{i}_y E_{0ty} + \mathbf{i}_z E_{0tz}) \exp\,[j(\omega t - \beta_{2x} x + \beta_{2z} z)] \tag{17.164}$$

where E_{0tx}, E_{0ty}, E_{0tz}, β_{2x}, β_{2z} are also unknown constants. Again there is no variation along the y-direction. The **H** vectors are represented similarly in which the letter E is replaced by H, and the origin of the coordinate system is chosen on the interface plane $z = 0$, as shown in Figure 17.6.

First, we evaluate the wave numbers. On the interface $z = 0$, the exponents must be equal for $\mathbf{E}_i$ and $\mathbf{E}_r$.

$$\therefore \quad \beta_{1x} = \beta_1 \sin\,\theta_i = \left(\frac{2\pi n_1}{\lambda_0} \right) \sin\,\theta_i \tag{17.165}$$

To evaluate β_{1z}, substituting in the wave equation in the medium 1,

$$\beta_{1x}^2 + \beta_{1z}^2 = \mu_0 \varepsilon_1 \omega^2 = \left(\frac{2\pi n_1}{\lambda_0} \right)^2 \tag{17.166}$$

$$\therefore \quad \beta_{1z}^2 = \left(\frac{2\pi n_1}{\lambda_0} \right)^2 (1 - \sin^2\theta_i)$$

$$\therefore \quad \beta_{1z} = \left(\frac{2\pi n_1}{\lambda_0} \right) \cos\,\theta_i = \beta_1 \cos\,\theta_i \tag{17.167}$$

$$\therefore \quad \mathbf{E}_r = \mathbf{E}_{0r} \exp\,[j\{\omega t - \beta_1(x \sin\,\theta_i - z \cos\,\theta_i)\}] \tag{17.168a}$$

which is a plane wave reflected from the interface at an angle equal to the angle of incidence.

FOR THE TRANSMITTED WAVE

Again equating the exponents of the incident and the transmitted waves on the interface $z = 0$, we have

$$\beta_{2x} = \beta_1 \sin \theta_i = \left(\frac{2\pi n_1}{\lambda_0}\right) \sin \theta_i \tag{17.168b}$$

To evaluate β_{2z}, using the wave equation (17.161) for $\mathbf{E}_t$

$$\beta_{2x}^2 + \beta_{2z}^2 = \omega^2 \mu_0 \varepsilon_2$$

$$\therefore \quad \beta_{2z} = \pm \left[\left(\frac{2\pi n_2}{\lambda_0}\right)^2 - \left(\frac{2\pi n_1}{\lambda_0}\right)^2 \sin^2 \theta_i \right]^{1/2}$$

$$= \pm j \left(\frac{2\pi n_2}{\lambda_0}\right) \left[\left(\frac{n_1}{n_2}\right)^2 \sin^2 \theta_i - 1 \right]^{1/2} \tag{17.169}$$

$\therefore$ The transmitted wave is

$$\mathbf{E}_t = \mathbf{E}_{0t} \exp \left[j \left\{ \omega t - \left(\frac{2\pi n_1}{\lambda_0}\right) x \sin \theta_i \right\} - \left(\frac{2\pi n_2}{\lambda_0}\right) \left\{ \left(\frac{n_1}{n_2}\right)^2 \sin^2 \theta_i - 1 \right\}^{1/2} z \right] \tag{17.170}$$

The sign before the z-term of the exponent is –ve, because $\mathbf{E}$ cannot become infinite as $z \to \infty$.

Note that

$$\beta_{2z} = j \left(\frac{2\pi n_2}{\lambda_0}\right) \left[\left(\frac{n_1}{n_2}\right)^2 \sin^2 \theta_i - 1 \right]^{1/2}$$

$$\therefore \quad \cos \theta_t = - \left[1 - \left(\frac{n_1}{n_2}\right)^2 \sin^2 \theta_i \right]^{1/2} = - \left(1 - \sin^2 \theta_t\right)^{1/2} \tag{17.171}$$

Next we evaluate the amplitudes of $\mathbf{E}$ and $\mathbf{H}$ in the reflected and the transmitted waves. We take the incident $\mathbf{E}$ wave normal to the plane of incidence. Hence the reflected and the transmitted $\mathbf{E}$ waves will also be normal to the plane of incidence, i.e.

$$\mathbf{E}_i = \mathbf{i}_y E_{0i} \exp [j(\omega t - \beta_1 \mathbf{n}_i \cdot \mathbf{r})] \tag{17.172a}$$

$$\mathbf{E}_r = \mathbf{i}_y E_{0r} \exp [j(\omega t - \beta_1 \mathbf{n}_r \cdot \mathbf{r})] \tag{17.172b}$$

$$\mathbf{E}_t = \mathbf{i}_y E_{0t} \exp [j(\omega t - \beta_{2x} x - \beta_{2z} z)] \tag{17.172c}$$

writing the exponents in a more compact form, $\mathbf{n}_i$ and $\mathbf{n}_r$ being the unit vectors along the ray of incidence and the ray of reflection respectively; and β_{2x} and β_{2z} are now known quantities.

For the $\mathbf{H}$ vector, $\mathbf{H}$ is now in the plane of incidence in the incident and the reflected waves, and will have both x- and z-components, i.e.

$$\mathbf{H}_i = H_{0i}(\mathbf{i}_x \cos \theta_i + \mathbf{i}_z \sin \theta_i) \exp [j(\omega t - \beta_1 \mathbf{n}_i \cdot \mathbf{r})] \tag{17.173a}$$

$$\mathbf{H}_r = H_{0r}(-\mathbf{i}_x \cos\theta_i + \mathbf{i}_z \sin\theta_i) \exp[j(\omega t - \beta_1 \mathbf{n}_r \cdot \mathbf{r})] \tag{17.173b}$$

For the transmitted wave, the amplitude expression is kept more general,

$$\mathbf{H}_t = (\mathbf{i}_x H_{0tx} + \mathbf{i}_y H_{0ty} + \mathbf{i}_z H_{0tz}) \exp[j(\omega t - \beta_{2x} x - \beta_{2z} z)] \tag{17.174}$$

Applying the boundary condition of continuous tangential **E** on the interface $z = 0$, we get

$$E_{0i} + E_{0r} = E_{0t} \tag{17.175}$$

The continuity of the tangential **H** on $z = 0$ requires

$$H_{0ty} = 0 \quad \text{and} \quad (H_{0i} - H_{0r})\cos\theta_i = H_{0tx}, \text{ which is}$$

$$\sqrt{\frac{\varepsilon_0}{\mu_0}}\; n_1(E_{0i} - E_{0r})\cos\theta_i = H_{0tx} \tag{17.176}$$

Continuity of normal **B** (or of **H**, since the media are non-magnetic) at the interface $z = 0$, gives

$$\sqrt{\frac{\varepsilon_0}{\mu_0}}\, n_1(E_{0i} + E_{0r})\sin\theta_i = H_{0tz} \tag{17.177}$$

So we have three equations, i.e. (17.175) to (17.177) to evaluate four unknowns E_{0r}, E_{0t}, H_{0tx}, H_{0tz}. Hence we choose one of the Maxwell's equations and apply it to the transmitted wave, i.e.

$$\nabla \cdot \mathbf{B}_t = \mu_0 \nabla \cdot \mathbf{H}_t = 0$$

$$\therefore \quad \beta_{2x} H_{0tx} + \beta_{2z} H_{0tz} = 0 \tag{17.178}$$

$\therefore$ Solving these equations, we get

$$\left(\frac{H_{0tx}}{E_{0i}}\right)_N = -\left(\sqrt{\frac{\varepsilon_0}{\mu_0}}\right) \frac{j\left[2n_1 \cos\theta_i \sqrt{\sin^2\theta_i - \left(\frac{n_2}{n_1}\right)^2}\right]}{\cos\theta_i - j\sqrt{\sin^2\theta_i - \left(\frac{n_2}{n_1}\right)^2}} \tag{17.179a}$$

$$\left(\frac{H_{0tz}}{E_{0i}}\right)_N = \frac{-\left(\sqrt{\frac{\varepsilon_0}{\mu_0}}\right)(n_1 \sin 2\theta_i)}{\cos\theta_i - j\sqrt{\sin^2\theta_i - \left(\frac{n_2}{n_1}\right)^2}} \tag{17.179b}$$

Thus for the transmitted wave, H_{0tx} is imaginary and H_{0tz} is real, and hence the two components are 90° out of phase; thus this wave rotates in the plane of incidence.

If the Poynting vector was evaluated, it will be seen that the z-component of $\bar{\mathbf{S}}_{av}$ is zero. Thus the behaviour of the wave patterns during the phenomenon of total reflection explains how an optical fibre is capable of *guiding a wave*, while offering *low* radiative losses.

17.16 REFLECTION AND REFRACTION AT THE SURFACE OF A GOOD CONDUCTOR

We have seen that the wave number of a good conductor (in Section 17.6) is

$$k_2 = \frac{2\pi n_2}{\lambda_0} = \sqrt{\frac{\omega\mu_2\sigma_2}{2}}\,(1 - j)$$

$$= \frac{1}{d\sqrt{2}}(1 - j) = \frac{1}{\delta}(1 - j) \tag{17.180}$$

where δ (or $d\sqrt{2}$) is the skin-depth. Since the index of refraction n_2 is complex and very large, the application of the Snell's law

$$\frac{\sin\theta_t}{\sin\theta_i} = \frac{n_1}{n_2}$$

leads to a very small value of sin θ_t. Thus,

$$\sin\theta_t = \frac{n_1}{n_2}\sin\theta_i \quad \text{and} \quad \cos\theta_t = \sqrt{1 - \left(\frac{n_1}{n_2}\right)^2 \sin^2\theta_i}$$

Though cos θ_t is complex, but because $n_2 >> n_1$, the imaginary part is negligible, and

$$\cos\theta_t \simeq 1 \rightarrow \theta_t \simeq 0 \rightarrow \sin\theta_t \simeq 0 \tag{17.181}$$

Thus we can justifiably say that the wave penetrates the conductor normally, whatever the angle of incidence might be. Hence the concept of the skin-depth applies to an electromagnetic wave incident at any angle to the surface of a good conductor. For all values of θ_i, the transmitted wave is a plane wave propagating along the normal to the surface, with enormous damping which is the characteristic of these waves in good conductors.

When the incident wave has its **E** vector normal to the plane of incidence, the Fresnel's equations are

$$\left(\frac{E_{0r}}{E_{0i}}\right)_N = \frac{n_1\mu_{r2}\cos\theta_i - \left(\frac{\lambda_0}{2\pi\delta}\right)(1 - j)}{n_1\mu_{r2}\cos\theta_i + \left(\frac{\lambda_0}{2\pi\delta}\right)(1 - j)} \simeq -1 \tag{17.182a}$$

$$\left(\frac{E_{0t}}{E_{0i}}\right)_N = \frac{2n_1\mu_{r2}\cos\theta_i}{n_1\mu_{r2}\cos\theta_i + \left(\frac{\lambda_0}{2\pi\delta}\right)(1 - j)} << 1 \tag{17.182b}$$

The first ratio of Eq. (17.182) implied that $\mathbf{E}_r$ of the reflected wave is in the opposite direction to that of $\mathbf{E}_{0i}$. The transmitted wave is a weak wave with a high degree of attenuation. For the reflected wave, there is a slight loss of intensity, whereas in the case of total reflection there was no loss of intensity and the reflection coefficient in that case was equal to unity.

When the *E* wave is parallel to the plane of incidence, the behaviour pattern is similar to that of the previous case, i.e.

$$\left(\frac{E_{0r}}{E_{0i}}\right)_P \simeq -1 \quad \text{and} \quad \left(\frac{E_{0t}}{E_{0i}}\right)_P << 1 \tag{17.183}$$

The wave components are shown diagrammatically in Figure 17.21.

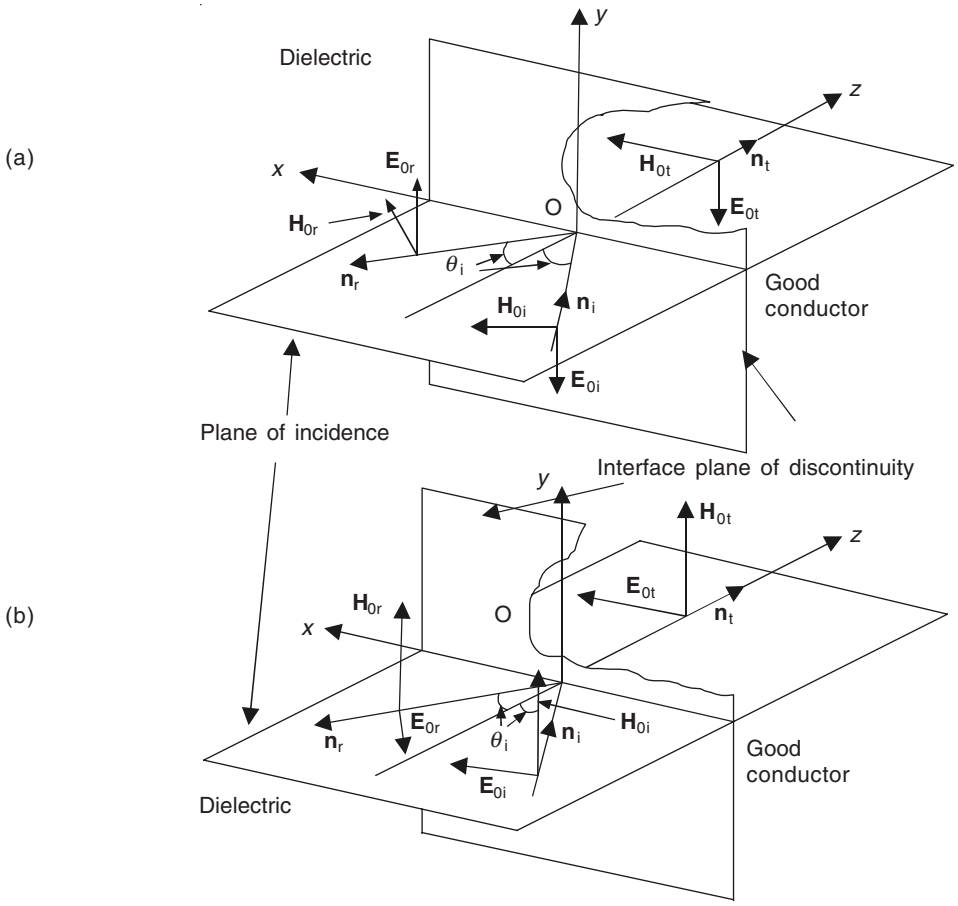


Figure 17.21 The incident, reflected and transmitted waves across the interface between a dielectric and a good conductor: (a) The incident wave with the **E** vector normal to the plane of incidence; (b) the incident wave with its **E** vector parallel to the plane of incidence. The coefficient of reflection is approximately unity in both the cases. The tangential **E** of the incident and the reflected waves nearly cancel each other on the interface.

17.17 OBLIQUE INCIDENCE OF UNIFORM PLANE WAVES ON A PERFECTLY CONDUCTING SURFACE

We shall close this chapter by considering this limiting boundary condition problem, which will serve as an introduction to the study of wave guiding in the next chapter. As in a number of previous problems, in this problem also, we shall study two cases, i.e.

1. Incident **E** wave polarized in the plane of incidence
2. Incident **E** wave polarized normal to the plane of incidence.

Note: In the next section, we shall establish the geometrical basis of the exponents of the expressions for the obliquely travelling waves (i.e. waves not travelling along the coordinate directions). In fact, this part of the analysis explains the physical basis of the mathematical expressions used from Sections 17.13 onwards.

17.17.1 Oblique Incidence of Uniform Plane E.M. Waves on a Perfectly Conducting Surface: E Wave Parallel to the Plane of Incidence

As before, we consider a ray normal to the wave-front, making an angle θ_i with the normal to the interface. The plane of incidence is the x-z plane (Figure 17.22).

Since energy cannot pass through a perfect conductor, only a reflected wave exists, whose ray makes an angle, say, θ_r with the normal to the interface.

The electric and the magnetic field vectors lie perpendicular to the respective directions of propagation (being the property of uniform plane waves). The **E** vector lies in the plane of the paper [Figure 17.22(b)], which is also the plane of incidence in that diagram; and the **H** vector is normal to the plane of the paper.

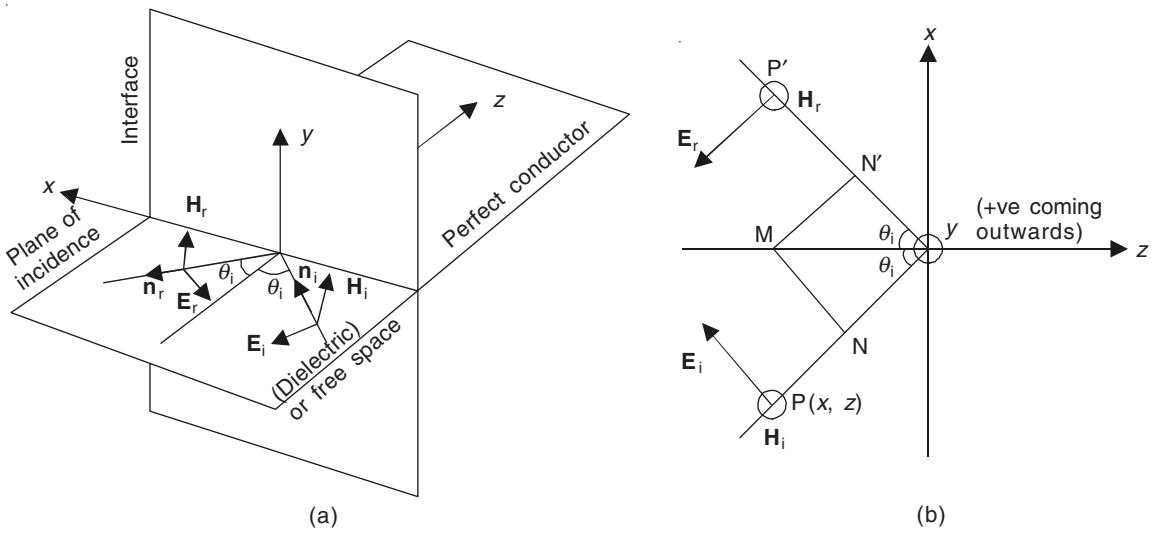


Figure 17.22 The incident and the reflected waves, for oblique incidence (with the **E** vector in the plane of incidence) on the perfectly conducting interface ($\sigma \to \infty$): (a) Isometric view; (b) view along the plane of incidence (z-x plane).

Let us say that a wave is now approaching in a direction PO making an angle θ_i with the normal to the interface. $\mathbf{E}_i$ lies in the plane z-x, and will be at right angles to PO as shown. $\mathbf{B}_i$ will be parallel to Oy. Let the point P on the ray of the wave be the point $(-x, 0, -z)$, then (as shown in the figure), PM is parallel to the x-axis and MN perpendicular to PO.

$$\therefore \qquad \text{PO} = \text{PN} + \text{NO} = z \cos \theta_i + x \sin \theta_i \; (= l) \tag{17.184}$$

Hence the factor $\exp(-j\beta z)$ (for the normally incident waves) is replaced by $\exp[-j\beta(z \cos \theta_i + x \sin \theta_i)]$, since z in the case of the normally incident wave and $(z \cos \theta_i + x \sin \theta_i)$, in this case, measures the distance along the line of wave propagation; and

$$\mathbf{E}_i = \mathbf{E}_{0i+} \exp j[\omega t - \beta (z \cos \theta_i + x \sin \theta_i)] \tag{17.185}$$

with which is associated the **H** wave

$$\mathbf{H}_i = \mathbf{H}_{0i+} \exp j[\omega t - \beta (z \cos \theta_i + x \sin \theta_i)] \tag{17.186}$$

in the direction shown in Figure 17.15 and Figure 17.22(b).

These waves, i.e. represented by Eqs. (17.185) and (17.186), could be described as a train of waves of amplitude

$$\mathbf{E}_{0i+} \exp(-j\beta x \sin\theta_i), \qquad \mathbf{H}_{0i+} \exp(-j\beta x \sin\theta_i)$$

moving towards the interface $z = 0$, with a velocity c sec θ_i. By comparing with the case of normal incidence (Section 17.9), these waves give rise to the reflected waves along OP′, where

$$\text{OP}' = l' = (z \cos\theta_r - x \sin\theta_r) \tag{17.187}$$

∴ The reflected waves are

$$\mathbf{E}_r = \mathbf{E}_{0r} \exp(+j\beta l') = \mathbf{E}_{0r-} \exp j[\omega t + \beta(-x \sin\theta_r + z \cos\theta_r)] \tag{17.188a}$$

$$\mathbf{H}_r = \mathbf{H}_{0r} \exp(+j\beta l') = \mathbf{H}_{0r-} \exp j[\omega t + \beta(-x \sin\theta_r + z \cos\theta_r)] \tag{17.188b}$$

∴ The resultant wave pattern due to the combination of the incident and the reflected wave is

$$E_x = [E_{0i+} \cos\theta_i \exp\{-j\beta(x \sin\theta_i + z \cos\theta_i)\} - E_{0r-} \cos\theta_r \exp\{j\beta(-x \sin\theta_r + z \cos\theta_r)\}] \exp(j\omega t) \tag{17.189a}$$

$$E_z = [-E_{0i+} \sin\theta_i \exp\{-j\beta(x \sin\theta_i + z \cos\theta_i)\} - E_{0r-} \sin\theta_r \exp\{j\beta(-x \sin\theta_r + z \cos\theta_r)\}] \exp(j\omega t) \tag{17.189b}$$

It should be noted that since **E** lies in the plane of incidence, it has two components, i.e. x- and z-, whilst **H** will be along the y-axis. Hence,

$$H_y = [H_{0i+} \exp\{-j\beta(x \sin\theta_i + z \cos\theta_i)\} + H_{0r-} \exp\{j\beta(-x \sin\theta_r + z \cos\theta_r)\}] \exp(j\omega t) \tag{17.190}$$

Now we apply the boundary conditions on the interface of the perfect conductor (i.e. $z = 0$ plane):
No tangential **E** on $z = 0$, i.e. $E_x = 0$, for all x

∴ $(E_x)_{z=0} = 0 = E_{0i+} \cos\theta_i \exp(-j\beta x \sin\theta_i) - E_{0r-} \cos\theta_r \exp(-j\beta x \sin\theta_r)$

∴ $\theta_r = \theta_i \rightarrow$ angle of reflection = angle of incidence, and $E_{0i+} = E_{0r-}$, i.e. the amplitudes of the incident and the reflected waves are the same. Also, the characteristic impedance relationship gives

$$\frac{E_{0i+}}{H_{0i+}} = Z_0 = \frac{E_{0r-}}{H_{0r-}} \qquad \text{(for free space)}$$

∴ Substituting in Eqs. (17.189) and (17.190), we get

$$\mathbf{E}_x = \mathbf{i}_x[-2jE_{0i+} \cos\theta_i \sin(\beta z \cos\theta_i) \exp j(\omega t - \beta x \sin\theta_i)] \tag{17.191a}$$

$$\mathbf{E}_z = \mathbf{i}_z[-2E_{0i+} \sin\theta_i \cos(\beta z \cos\theta_i) \exp j(\omega t - \beta x \sin\theta_i)] \tag{17.191b}$$

$$\mathbf{H}_y = \mathbf{i}_y\left[2\left(\frac{E_{0i+}}{Z_0}\right)\cos(\beta z \cos\theta_i) \exp j(\omega t - \beta x \sin\theta_i)\right] \tag{17.191c}$$

These equations represent travelling waves with respect to the x-direction, but standing waves with respect to z-direction.

On the perfectly conducting plane $z = 0$, $\mathbf{E}_x = 0$ at all instants of time; and also in parallel planes at distances (nd) in front of the perfect conductor surface $z = 0$, where

$$z = d = \frac{\pi}{\beta \cos \theta_i} = \frac{\lambda}{2 \cos \theta_i} = \frac{1}{2f\sqrt{\mu_0 \varepsilon_0} \cos \theta_i} \tag{17.192}$$

$$\left\{ \because \qquad \beta = \frac{2\pi}{\lambda} \qquad \text{and} \qquad \beta z \cos \theta_i = \pi \Rightarrow z = \frac{\pi}{\beta \cos \theta_i} \right\}$$

The alternating amplitude of $\mathbf{E}_x$ is maximum in the planes which are odd multiples of (d/2) in front of the conducting interface $z = 0$.

H_y and E_z are maximum, when $\mathbf{E}_x = 0$, i.e. $z = 0$, and (nd), n being integers.

H_y and E_z are zero, when $\mathbf{E}_x$ = maximum, i.e. z = odd multiple of (d/2).

Also H_y and E_z are 90° (= π/2) out of phase with $\mathbf{E}_x$.

It should also be noted that the distance between the successive maxima and the zero, measured normal to the interface plane becomes greater as the incidence becomes more oblique, i.e. as θ_i increases.

17.17.2 Oblique Incidence of Uniform Plane E.M. Waves on a Perfectly Conducting Surface: E Vector Normal to the Plane of Incidence

As before in the previous section, we consider again a wave approaching in a direction PO which makes an angle θ_i with the normal to the interface of discontinuity. The **E** vector is now parallel to Oy (the origin of the coordinate system being on the interface plane as shown in Figure 17.23); and hence in Figure 17.23(b) it is normal to the plane of the paper, but **B** or **H** will lie in the plane of incidence (Ozx), and will be at right angles to PO, as shown. Let P be the point ($-x$, 0, $-z$), then PM is parallel to the x-axis and MN is perpendicular to PO, and hence:

$$\text{PO} = \text{PN} + \text{NO} = z \cos \theta_i + x \sin \theta_i \; (= 1) \tag{17.193}$$

Hence, comparing with the normal incidence problem, the factor exp ($-j\beta z$) is now replaced by exp [$-j\beta$ (z cos θ_i + x sin θ_i)], since z in the previous case (Section 17.9) and (z cos θ_i + x sin θ_i) in the present case measure distances along the line of propagation; and

$$\begin{aligned} \mathbf{E}_i &= \mathbf{E}_{0i} \exp [-j\beta (z \cos \theta_i + x \sin \theta_i)] \\ &= \mathbf{E}_{0i+} \exp j [\omega t - \beta (z \cos \theta_i + x \sin \theta_i)] \end{aligned} \tag{17.194}$$

with which is associated the **H** wave;

$$\begin{aligned} \mathbf{H}_i &= \mathbf{H}_{0i} \exp [-j\beta (z \cos \theta_i + x \sin \theta_i)] \\ &= \mathbf{H}_{0i+} \exp j [\omega t - \beta (z \cos \theta_i + x \sin \theta_i)] \end{aligned} \tag{17.195}$$

in the directions shown in Figure 17.23.

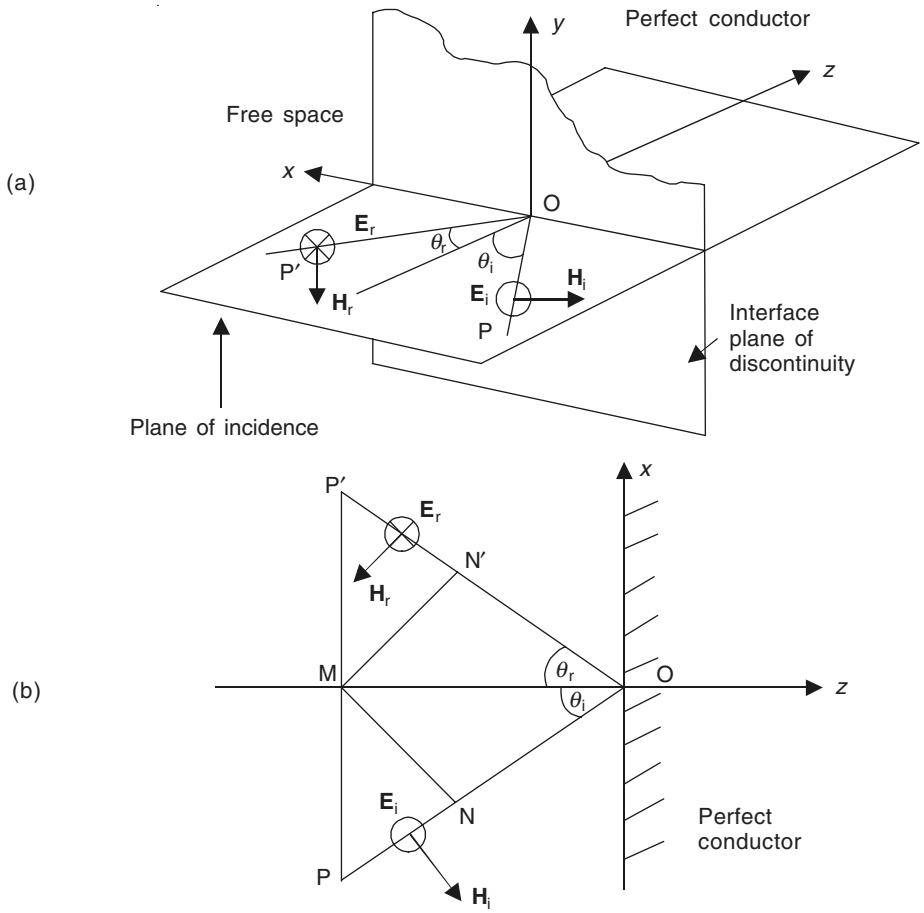


Figure 17.23 The incident and the reflected waves, for oblique incidence, with **E** vector normal to the plane of incidence, on a perfectly conducting interface ($\sigma \to \infty$): (a) Isometric view; and (b) view along the plane of incidence.

The previous two equations, i.e. (17.194) and (17.195) could also be described as a train of waves of amplitude $\mathbf{E}_{0i} \exp(-j\beta x \sin \theta_i)$ and $\mathbf{H}_{0i} \exp(-j\beta x \sin \theta_i)$ moving towards the boundary interface $z = 0$ with a velocity (c sec θ_i). By the Section 17.9, these give rise to reflected waves in which the signs of $\mathbf{E}_{0i}$ (but not $\mathbf{H}_{0i}$) and of z are changed, i.e.

$$\mathbf{E}_r = -\mathbf{E}_{0i} \exp[j\beta(z \cos \theta_i - x \sin \theta_i)] \tag{17.196a}$$

$$\mathbf{H}_r = -\mathbf{H}_{0i} \exp[j\beta(z \cos \theta_i - x \sin \theta_i)] \tag{17.196b}$$

We have written down the expressions for the reflected waves, based on physical arguments. However, alternatively, the same results can be obtained by using formal mathematics based on the Maxwell's equations, which we indicate below.

The ray OP, in Figure 17.23, is normal to the wave-front, and makes an angle θ_i with the normal to the interface of the conducting region, i.e. the plane $z = 0$. Since there is no energy penetration through a perfect conductor, only the reflected wave exists, which, say, makes an angle θ_r with the normal to the interface plane.

Both **E** and **H** are normal to the direction of propagation, **E** being normal to the plane of the paper, and **H** lying in the plane of the paper [referring to Figure 17.23(b)].

$\mathbf{E}_i$ and $\mathbf{E}_r$ are the incident and the reflected electric waves respectively, and $\mathbf{H}_i$ and $\mathbf{H}_r$ are the incident and the reflected magnetic waves respectively. ($\mathbf{E} \times \mathbf{H}$) would give the direction of propagation for each wave.

We consider the incident wave first. On the ray, at a point P, the length OP—along the direction of propagation,

$$\begin{aligned} \text{OP} &= l = (x \sin \theta_i + z \cos \theta_i) \quad [\text{(Figure 17.23(b)}] \\ \mathbf{E}_i &= \mathbf{i}_y E_{0i} \exp(-j\beta l) = \mathbf{i}_y E_{0i} \exp[-j\beta(x \sin \theta_i + z \cos \theta_i)] \\ &= \mathbf{i}_y E_{0i+} \exp[j\{\omega t - \beta(x \sin \theta_i + z \cos \theta_i)\}] \end{aligned} \tag{17.197}$$

To find the corresponding **H** field, we use the Maxwell's equation,

$$\nabla \times \mathbf{E} = -\left(\frac{\partial \mathbf{B}}{\partial t}\right)$$

or

$$\mathbf{i}_x \left[-\left(\frac{\partial \mathbf{E}_y}{\partial z}\right)\right] + \mathbf{i}_y 0 + \mathbf{i}_z \frac{\partial \mathbf{E}_y}{\partial x} = -j\omega\mu_0(\mathbf{i}_x H_x + \mathbf{i}_y 0 + \mathbf{i}_z H_z)$$

since there is only the y-component of **E**, and there is no variation in the y-direction.

$$\therefore \quad -\left(\frac{\partial E_y}{\partial z}\right) = -j\omega\mu_0 \mathbf{i}_x H_x \quad \text{and} \quad \left(\frac{\partial E_y}{\partial x}\right) = -j\omega\mu_0 H_z$$

Now $\beta = \omega\sqrt{\mu_0\varepsilon_0}$

$$\therefore \quad \frac{\beta}{\omega\mu_0} = \sqrt{\frac{\varepsilon_0}{\mu_0}} = \frac{1}{Z_0}$$

$$\therefore \quad H_{ix} = \left(\frac{1}{j\omega\mu_0}\right)\left(\frac{\partial E_y}{\partial z}\right)$$

$$= -\left[\left(\frac{j\beta\cos\theta_i}{j\omega\mu_0}\right)\right] E_{0i} \exp[-j\beta(x \sin \theta_i + z\cos\theta_i)]$$

$$= -\left(\frac{E_{0i}}{Z_0}\right)\cos\theta_i \exp[-j\beta(x \sin \theta_i + z\cos\theta_i)] \tag{17.198}$$

$$H_{iz} = -\left(\frac{1}{j\omega\mu_0}\right)\left(\frac{\partial E_y}{\partial x}\right)$$

$$= \left(\frac{j\beta \sin \theta_i}{j\omega\mu_0}\right) E_{0i} \exp[-j\beta(x \sin \theta_i + z \cos \theta_i)]$$

$$= \left(\frac{E_{0i}}{Z_0}\right) \sin \theta_i \exp[-j\beta(x \sin \theta_i + z \cos \theta_i)] \tag{17.199}$$

Next, we derive the expressions for the reflected wave,

$$\mathbf{E}_r = \mathbf{i}_y E_{0r} \exp(-j\beta l'), \qquad \text{where } l' = OP' = (-x \sin\theta_r + z\cos\theta_r),$$

where the –ve sign is due to θ_r being measured in the sense opposite to that of θ_i.

$$\therefore \qquad \mathbf{E}_r = \mathbf{i}_y E_{0r} \exp[j\beta(-x\sin\theta_r + z\cos\theta_r)] \tag{17.200}$$

and the corresponding **H** vector:

$$H_{rx} = \left(\frac{1}{j\omega\mu_0}\right)\left(\frac{\partial E_y}{\partial z}\right)$$

$$= \left(\frac{j\beta\cos\theta_r}{j\omega\mu_0}\right) E_{0r} \exp[j\beta(-x\sin\theta_r + z\cos\theta_r)]$$

$$= \left(\frac{E_{0r}}{Z_0}\right)\cos\theta_r \exp[j\beta(-x\sin\theta_r + z\cos\theta_r)] \tag{17.201}$$

$$H_{rz} = -\left(\frac{1}{j\omega\mu_0}\right)\left(\frac{\partial E_y}{\partial x}\right)$$

$$= -\left(\frac{-j\beta\sin\theta_r}{j\omega\mu_0}\right) E_{0r} \exp[j\beta(-x\sin\theta_r + z\cos\theta_r)]$$

$$= \left(\frac{E_{0r}}{Z_0}\right)\sin\theta_r \exp[j\beta(-x\sin\theta_r + z\cos\theta_r)] \tag{17.202}$$

$\therefore$ Combining the incident and the reflected waves, the resultant is obtained as

$$E_y = E_{0i}\exp[-j\beta(x\sin\theta_i + z\cos\theta_i)] + E_{0r}\exp[j\beta(-x\sin\theta_r + z\cos\theta_r)] \tag{17.203a}$$

$$H_x = -\left(\frac{E_{0i}}{Z_0}\right)\exp[-j\beta(x\sin\theta_i + z\cos\theta_i)]\cos\theta_i + \left(\frac{E_{0r}}{Z_0}\right)\exp[j\beta(-x\sin\theta_r + z\cos\theta_r)]\cos\theta_r \tag{17.203b}$$

$$H_z = \left(\frac{E_{0i}}{Z_0}\right)\exp[-j\beta(x\sin\theta_i + z\cos\theta_i)]\sin\theta_i + \left(\frac{E_{0r}}{Z_0}\right)\exp[j\beta(-x\sin\theta_r + z\cos\theta_r)]\sin\theta_r \tag{17.203c}$$

Now we apply the boundary condition on the reflecting plane $z = 0$ which is that there is no tangential **E** on the perfectly conducting surface, i.e.

for $z = 0$, $E_y = 0$

$$= E_{0i}\exp[-j\beta(x\sin\theta_i)] + E_{0r}\exp[-j\beta(x\sin\theta_r)], \text{ for all } x \text{ (and } t)$$

$$\therefore \quad \theta_i = \theta_r \qquad \text{and} \qquad E_{0r} = -E_{0i}$$

$\therefore$ The resultant field components are

$$E_y = -j2E_{0i} \sin(\beta z \cos\theta_i) \exp(-j\beta x \sin\theta_i)$$
$$= -j2E_{0i+} \sin(\beta z \cos\theta_i) \exp j(\omega t - \beta x \sin\theta_i) \quad (17.204a)$$

$$H_x = -2\left(\frac{E_{0i}}{Z_0}\right)\cos\theta_i \cos(\beta z \cos\theta_i) \exp(-j\beta x \sin\theta_i)$$
$$= -2\left(\frac{E_{0i+}}{Z_0}\right)\cos\theta_i \cos(\beta z \cos\theta_i) \exp j(\omega t - \beta x \sin\theta_i) \quad (17.204b)$$

$$H_z = -2j\left(\frac{E_{0i}}{Z_0}\right)\sin\theta_i \sin(\beta z \cos\theta_i) \exp(-j\beta x \sin\theta_i)$$
$$= -2j\left(\frac{E_{0i+}}{Z_0}\right)\sin\theta_i \sin(\beta z \cos\theta_i) \exp j(\omega t - \beta x \sin\theta_i) \quad (17.204c)$$

$\therefore$ These are travelling waves in the x-direction, but standing patterns in the z-direction.

On the plane $z = 0$ (i.e. the interface between the perfect conductor and the free space), $E_y = 0$, $H_z = 0$, i.e. no tangential **E** field and no normal **H** field; and only the tangential **H** (i.e. H_x) field exists.

$\therefore$ There will be circulating currents (i.e. eddy currents) in the y-direction, i.e. J_y, on this plane.

$\therefore$ On the plane $z = 0$, there will be zeroes of E_y and H_z; and maxima of H_x. These values will be repeated at intervals of $z = d$, such that $\sin(\beta d \cos\theta_i) = 0 = \sin\pi$ or $\sin n\pi$, where n = all integers.

$$\therefore \quad d_n = \frac{n\pi}{\beta\cos\theta_i} = \frac{n}{2f\sqrt{(\mu_0\varepsilon_0)}\cos\theta_i}, \quad \beta = 2\pi f\sqrt{\mu_0\varepsilon_0}$$

$$\therefore \quad d_1 = \frac{\pi}{\beta\cos\theta_i} \quad (17.205)$$

Also, E_y and H_z will have maxima, and H_x be zero at odd multiples of $(d_1/2)$ in front of the interface $z = 0$.

Hence d_1 = distance between successive zeroes and successive maxima; and $(d_1/2)$ = distance between successive zero and maximum.

Also, since $d_1 = \pi/(\beta \cos\theta_i)$, d_1 depends on θ_i—the angle of incidence of the wave on the reflecting surface. As θ_i increases to $(\pi/2)$, $\cos\theta_i \to 0$, and hence d_1 increases. A pattern of **E** wave crests (maxima) and zeroes is shown in Figure 17.24.

The standing wave pattern in the z-direction has been produced by the interaction between the incident wave and the reflected wave. Physically what has happened is when a positive crest of the incident **E** wave meets the reflecting surface ($z = 0$), it changes to a negative crest, and vice versa. The **H** and the **E** waves behave like water waves approaching a wall; and where the crests of the incident and the reflected waves coincide, at those points there is doubling of amplitudes. On the boundary (i.e. $z = 0$), whilst the resulting E_y is zero, the H_x has doubled because for H_x the positive crest is reflected as positive crest, and on the other hand H_z behaves like E_y wave. Thus, there are planes like A_1A_1', AA' (referring to Figure 17.24) which are in the same condition as the reflecting surface $z = 0$, i.e. on all these planes E_y and H_z are both zeroes and H_x maximum. The wave pattern is stationary in the z-direction, but gliding or travelling along the x-direction. Thus, *we have now 'guided' the wave along the x-direction.*

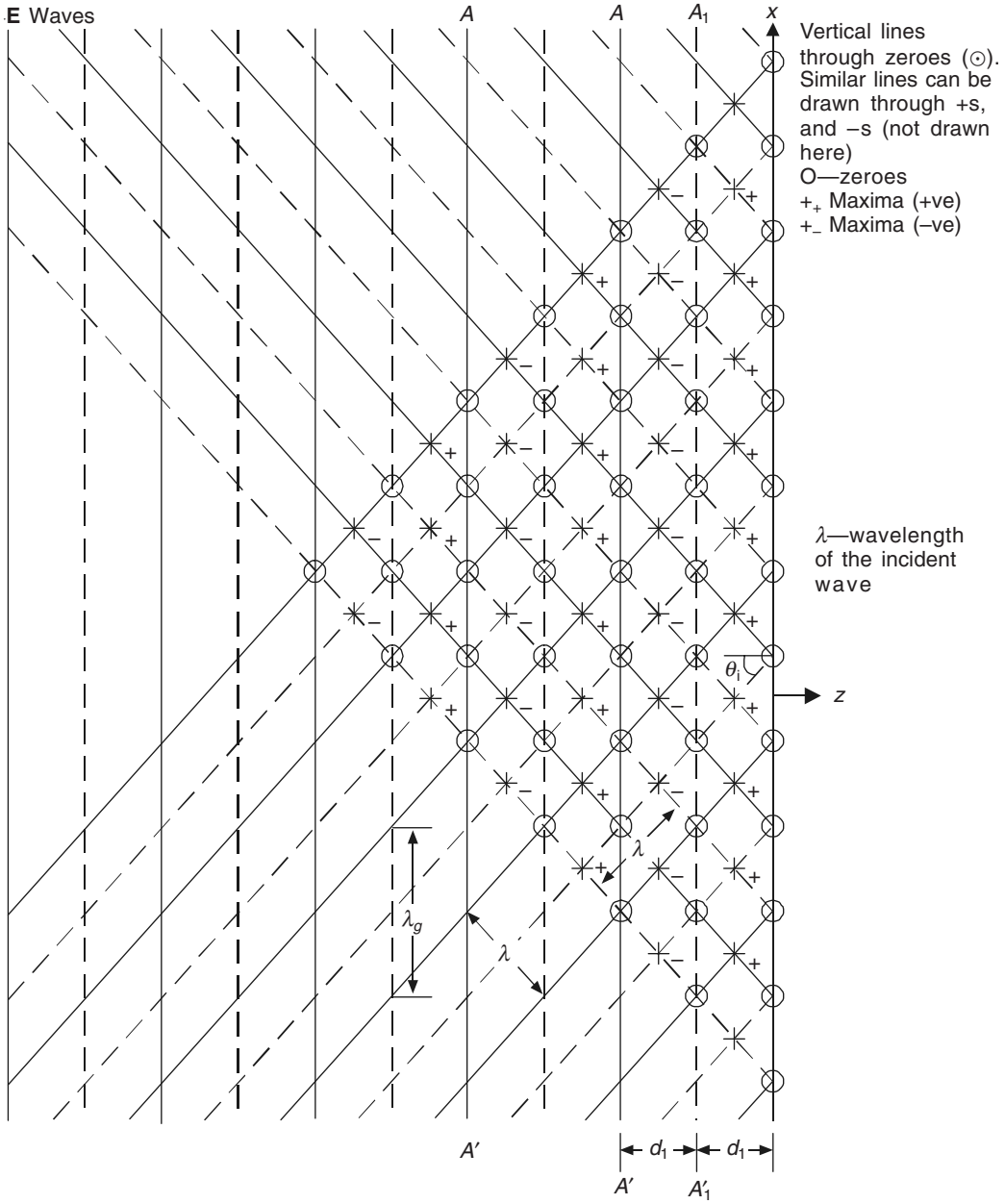


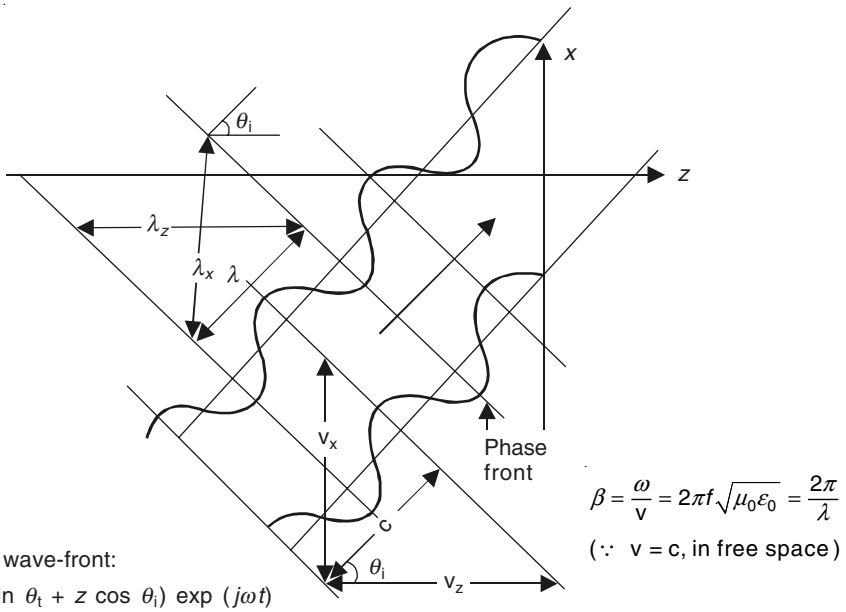
Figure 17.24 Resultant **E** wave pattern, when the incident **E** wave, normal to the plane of incidence, meets a perfectly conducting surface obliquely.

If λ is the wavelength of the incident wave (in the direction of its propagation), where $\lambda = 2\pi/\beta$, then the wavelength of the standing wave pattern in the z-direction (= λ_z) is given by

$$\lambda_z = 2d_1 = \frac{2\pi}{\beta \cos\theta_i} = \frac{\lambda}{\cos\theta_i} \tag{17.206}$$

and the wavelength of the guided wave along the x-direction (= λ_x) is given by

$$\lambda_x = \lambda_g = \frac{\lambda}{\sin \theta_i} \qquad \text{(also Figure 17.25)} \tag{17.207}$$



Obliquely propagating wave-front:

$E_i = \mathbf{i}_y E_{0i+} \exp [j\beta(x \sin \theta_t + z \cos \theta_i) \exp (j\omega t)$

$\lambda_z = \lambda/\cos \theta_i$; $\lambda_x = \lambda/\sin \theta_i$

$\beta_z = \beta \cos \theta_i$; $\beta_x = \beta \sin \theta_i$

The wave-front has travelled a distance c or u in one second.

$v_z = \omega/\beta_z = \omega/(\beta \cos \theta_i) = f\lambda_z = f\lambda/(\cos \theta_i)$

$v_x = \omega/\beta_x = \omega/(\beta \sin \theta_i) = f\lambda_x = f\lambda/(\sin \theta_i)$

Figure 17.25 An obliquely propagating plane wave, meeting a perfectly conducting surface, and its guided wave parameters.

Since the **E** vector is zero on every $z = (nd)$ (n being integers) plane away from the reflecting surface, if we put a conducting sheet (say) along the $z = d_1$ plane, then its presence will not affect the wave pattern between the planes $z = 0$ and $z = -d_1$; and we would now have a region between these two planes, where the waves are now 'guided' to travel along the x-direction (this technique forms the basis for the design of 'waveguides').

From the wavelengths obtained in Eqs. (17.205) and (17.206), the corresponding wave-numbers are

$$\beta_z = \text{the wave-number for the standing wave pattern} = \beta \cos \theta_i \tag{17.208a}$$

$$\beta_x = \text{the wave-number for the guided wave pattern} \quad = \beta \sin \theta_i \tag{17.208b}$$

$$\therefore \qquad \beta_z^2 + \beta_x^2 = \beta^2 \tag{17.209a}$$

and

$$\frac{1}{\lambda_z^2} + \frac{1}{\lambda_x^2} = \frac{1}{\lambda^2} \tag{17.209b}$$

Generalizing the position of the second conducting sheet (i.e. instead of keeping it along the plane $z = -d_1$), it can be placed on any of the zero-field planes without altering the wave pattern

between the planes $z = 0$ and $z = (md_1)$ where m is an integer. Such an operation is called an 'm-mode operation', whereas $m = 1$ is called 'mode 1 operation', for which,

$$2d_1 = \frac{2\pi}{\beta \cos \theta_i} = \frac{\lambda}{\cos \theta_i}$$

$\therefore$ $\cos \theta_1 = \dfrac{\lambda}{2d_1}$, and, in general, for mode m operation,

$$\cos \theta_m = \frac{m\lambda}{2d_1} \tag{17.210}$$

and, for this mode, the wavelength of the guided wave is

$$(\lambda_g =)\ \lambda_{gm} = \frac{\lambda}{\sin \theta_m} = \frac{\lambda}{\left[1 - \left(\dfrac{m\lambda}{2d_1}\right)^2\right]^{1/2}} \tag{17.211}$$

$\therefore$ The velocity of the guided wave (along the x-direction) is

$$\mathrm{v}_g = f\lambda_g = \frac{f\lambda}{\left[1 - \left(\dfrac{m\lambda}{2d_1}\right)^2\right]^{1/2}} = \mathrm{v}_x \tag{17.212}$$

where

$f\lambda$ = the velocity of the incident wave
= c, in free space
= v, in general, in any medium

The phase coefficient of the guided wave is

$$\beta_x = \beta_g = \frac{2\pi}{\lambda_x} = \frac{2\pi}{\lambda_g} = \left(\frac{2\pi}{\lambda}\right)\left[1 - \left(\frac{m\lambda}{2d_1}\right)^2\right]^{1/2}$$
$$= \beta \sin \theta_m \tag{17.213}$$

So, for the wave to propagate in this direction, β_x (or β_g) must be real, i.e. $E_{gx} \exp [j(\omega t - \beta_x x)]$.

$\therefore$ As λ is increased (or f is decreased), a point is reached at which β_x is zero, i.e. $\beta_x = \beta_g = 0$. If λ is increased still further, β_g becomes an imaginary quantity, and then there is no wave motion between the conducting plates; i.e. the frequency ($= f_c$) at which β_g is zero, is called the 'cut-off frequency' of the system. This gives

$$\beta_g = 0 \quad \text{or} \quad 1 - \left(\frac{m\lambda_c}{2d_1}\right)^2 = 0$$

$$\therefore \qquad \lambda_c = \frac{2d_1}{m} \tag{17.214}$$

$$\therefore \qquad f_c = \frac{\mathrm{v}}{\lambda_c} = \frac{\mathrm{v}m}{2d_1} \tag{17.215}$$

Given the order of the mode, m, the spacing d_1 and the velocity (to be general, in a dielectric) v, the cut-off frequency can be found.

∴ The phase velocity of the guided wave expressed in terms of the cut-off frequency is given as

$$v_g = \frac{v}{\left[1-\left(\frac{m\lambda}{2d_1}\right)^2\right]^{1/2}} = \frac{v}{\left[1-\left(\frac{m}{2d_1}\right)^2\left(\frac{v}{f}\right)^2\right]^{1/2}}$$

$$= \frac{v}{\left[1-\left(\frac{f_c}{f}\right)^2\right]^{1/2}} = v_x \qquad (17.216)$$

$$\therefore \quad \lambda_x \text{ or } \lambda_g = \frac{v_x}{f} = \frac{\lambda}{\left[1-\left(\frac{f_c}{f}\right)^2\right]^{1/2}} \qquad (17.217)$$

when $m = 1$, there is the longest wavelength and the lowest frequency. This is usable for a given guide frequency and is the mode most commonly used in rectangular waveguides, which are made by introducing two more parallel conducting plates at right angles to the direction of the **E** vector, and parallel to the z-axis. This is the basic structure of the rectangular waveguides (which we shall study in some further details in Chapter 18).

If f is >> f_c, then $v_g \doteq v$, and the wave-fronts in the guides are almost normal to the sides.

Also, note that the phase velocity of the guided wave $v_g > v$ [Eq. (17.26)] which in free space is the velocity of light (v = c). Thus though the 'wave-front' can travel at speeds greater than the velocity of light, any information can only pass down the guide at the speed with which any individual wave-front progresses in the x-direction, i.e. v sin θ_i, which is < v. To understand this point, it should be noted that the energy propagates in a direction normal to the wave-front with the velocity c or v (to be more general). For the guided wave travelling in the x-direction, the information or the energy will travel at a speed

$$v \cos\left(\frac{\pi}{2} - \theta_i\right) = v \sin \theta_i$$

This is the 'group velocity' of the wave, i.e.

$$v_{gr} = \left[1-\left(\frac{f_c}{f}\right)^2\right]^{1/2} \qquad (17.218)$$

this is due to the zigzag motion of the wave-front.

We shall now evaluate the Poynting vector for the resultant guided wave, i.e. from Eqs. (17.204), E_y with H_z gives a Poynting vector in the x-direction

$$\overline{S}_{x\,av} = \left(\frac{1}{2}\right) \text{Re}\left[\{-j2E_{0i+} \sin(\beta z \cos\theta_i) \exp(-j\beta x \sin\theta_i)\}\right.$$

$$\left. \times \left\{-2j\left(\frac{E_{0i+}}{Z_0}\right) \sin\theta_i \sin(\beta z \cos\theta_i) \exp(-j\beta x \sin\theta_i)\right\}\right]$$

$$= 2\left(\frac{E_{0i+}^2}{Z_0}\right)\sin\theta_i \sin^2(\beta z \cos\theta_i) \tag{17.219}$$

From the same equations, E_y with H_x would give a Poynting vector in the z-direction,

$$\bar{S}_{x\,av} = \left(\frac{1}{2}\right)\mathrm{Re}\left[-j2E_{0i+}\sin(\beta z\cos\theta_i)\exp(-j\beta x\sin\theta_i)\right]$$
$$\times\left[-2\left(\frac{E_{0i+}}{Z_0}\right)\cos\theta_i\cos(\beta z\cos\theta_i)\exp(-j\beta x\sin\theta_i)\right]$$
$$= 0, \text{ on averaging} \tag{17.220}$$

Thus it is seen that the energy flow is entirely parallel to the interface plane $z = 0$, along the x-axis.

In the system we have studied, it is seen that the **E** wave is at right angles to the direction of propagation (i.e. transverse to the direction of propagation) whereas the **H** wave has both the transverse and the longitudinal components. Hence this type of wave is called the TE wave (Transverse Electric); or some authors refer to it as the **H** wave. There are other modes of operation, such as TM (Transverse Magnetic) in which the magnetic field has only the transverse component whilst the **E** wave has both the transverse and the longitudinal components. There is also the TEM wave in which both the electric and the magnetic fields are transverse to the direction of propagation. Also, when the parallel conducting plates are positioned for mode 1 operation (i.e. in two successive zero planes), the pattern obtained is denoted as TE 1 mode. Higher modes are also possible. All these aspects are the subject matter of our study in Chapter 18.

17.18 PLANE WAVES ON CYLINDRICAL CONDUCTORS

So far we have seen that the plane waves can be propagated in free space and infinite dielectric media with plane faces. However plane waves can also be propagated along the z-direction (say, this direction being the direction of the axis) along a set of conductors, so long as all the sections of the conductor set normal to the z-axis are identical, i.e. the conductors are circular cylindrical, for example. Initially we assume the conductors to be perfect $(\sigma \to \infty \text{ or } \rho \to 0)$ and the medium to be loss-less. It follows from our study of skin effect and eddy currents in Chapter 15 that on perfect conductors the currents flow in 'infinitely thin surface layers' with no fields inside and there would be no energy dissipation. This problem can be posed in a number of ways.

In this case, the vector potential **A** as well as the Hertz vector $\mathbf{Z}_e$ would have only the z-component, i.e. parallel to the direction of the currents. The solution of the propagation equation for the Hertz vector would yield both a scalar and a vector potential (Section 13.9). By a suitable choice of the constraint equation, the scalar potential can be eliminated (i.e. Eq. 13.117), in which case the vector potential becomes normal to z-direction.

It is also possible to obtain the same result more directly by solving the scalar wave propagation equation, which is

$$\left(\frac{\partial^2 W}{\partial x^2} + \frac{\partial^2 W}{\partial y^2}\right) + \left(\frac{\partial^2 W}{\partial z^2} - \mu\varepsilon\frac{\partial^2 W}{\partial t^2}\right) \tag{17.221}$$

(for the loss-less dielectric medium)

It should be noted that even though the equation is written in Cartesian coordinate system, the cylindrical coordinate values can be obtained easily since $x = r \cos \phi$ and $y = r \sin \phi$. The solution to the above equation can be obtained by dividing the equation into two bracketted parts and setting each bracket to zero, i.e.

$$\frac{\partial^2 W}{\partial x^2} + \frac{\partial^2 W}{\partial y^2} = 0 \tag{17.222a}$$

and

$$\frac{\partial^2 W}{\partial z^2} - \mu\varepsilon \frac{\partial^2 W}{\partial t^2} = 0 \tag{17.222b}$$

The first equation is a two-dimensional Laplace's equation in x and y and the second equation is a one-dimensional wave equation in z and t.

The solution to the two-dimensional Laplace's equation can be written in terms of conjugate functions of complex variables. So if V_1 and V_2 [where $V_1 = V_1(x, y)$ and $V_2 = V_2(x, y)$] are the solutions of Eq. (17.222a), then W_1 and W_2 are the complex potential functions so that

$$W_1 = U_1 + jV_1 = F_1(x + jy) \quad \text{and} \quad W_2 = U_2 + jV_2 = F_2(x + jy) \tag{17.223}$$

where $j = \sqrt{-1}$.

Differentiating W_i with respect to x and y gives

$$\left.\begin{aligned} \frac{\partial W_i}{\partial x} &= \frac{\partial U_i}{\partial x} + j\frac{\partial V_i}{\partial x} = F_i'(x + jy) = \left\{ F_i'(z)\frac{\partial z}{\partial x} = F_i'(z) \right\} \\ \text{and} \\ \frac{\partial W_i}{\partial y} &= \frac{\partial U_i}{\partial y} + j\frac{\partial V_i}{\partial y} = jF_i'(x + jy) = \left\{ F_i'(z)\frac{\partial z}{\partial y} = jF_i'(z) \right\} \end{aligned}\right\} \tag{17.224}$$

In the above two equations, $z = x + jy$ and this z is not to be confused with the space variable z of Eqs. (17.221) and (17.222b).

From these two equations, equating the real and imaginary parts separately, we get

$$\frac{\partial V_i}{\partial x} = -\frac{\partial U_i}{\partial y} \quad \text{and} \quad \frac{\partial V_i}{\partial y} = \frac{\partial U_i}{\partial x} \tag{17.225}$$

i.e. the two sets (or families) of curves $U(x, y)$ = constant and $V(x, y)$ = constant intersect each other orthogonally. So one set may be chosen to be equipotential (so that this will be the potential function), and the other set would represent the lines of force (known as the stream function). From these the electric field intensity and the electric flux can be obtained as

$$\frac{dW_i}{dz} = \frac{dU_i + jdV_i}{dx + jdy} = \frac{\dfrac{\partial U_i}{\partial x}dx + \dfrac{\partial U_i}{\partial y}dy + j\left(\dfrac{\partial V_i}{\partial x}dx + \dfrac{\partial V_i}{\partial y}dy\right)}{dx + jdy}$$

$$= \frac{\dfrac{\partial V}{\partial y}(dx + jdy) + j\dfrac{\partial V}{\partial x}(dx + jdy)}{dx + jdy}$$

$$= \frac{\partial V}{\partial y} + j\frac{\partial V}{\partial x} = \frac{\partial U}{\partial x} - j\frac{\partial U}{\partial y} \tag{17.226}$$

by substituting from Eqs. (17.225).

Irrespective of whether U_i or V_i is the potential function, the magnitude of the electric field at a point would be given by $\left|\frac{\partial W}{\partial z}\right|$.

Let

dn be an element of length in the direction of maximum increase of potential, and

ds be the element of length obtained by rotating *dn* counterclockwise through $\pi/2$ radians, then

$$\left|\frac{dW}{dz}\right| = \frac{\partial U_i}{\partial n} = \frac{\partial V_i}{\partial s} \quad \text{or} \quad \left|\frac{dW}{dz}\right| = \frac{\partial V_i}{\partial n} = -\frac{\partial U_i}{\partial s} \tag{17.227}$$

depending on whether U or V is the potential function.

If V is the potential function, then the flux through any section of an equipotential surface is given by

$$\text{Flux} = -\varepsilon \int_{U_1}^{U_2} \frac{\partial V}{\partial n}\, ds = \varepsilon \int_{U_1}^{U_2} \frac{\partial U}{\partial s}\, ds = \varepsilon(U_2 - U_1) \tag{17.228}$$

Hence the capacitance per unit length is given by

$$C = \frac{|Q|}{|V_2 - V_1|} = \frac{\varepsilon |U|}{|V_2 - V_1|} \tag{17.229}$$

The solutions to the one-dimensional wave Eq. (17.222b) can be written as

$$f_1\left(z - t/\sqrt{\mu\varepsilon}\right) \qquad \text{and} \qquad f_2\left(z + t/\sqrt{\mu\varepsilon}\right) \tag{17.230}$$

Hence the solution to the scalar wave propagation Eq. (17.221) can be expressed as

$$W = V_1(x, y)\, f_1\left(z - t/\sqrt{\mu\varepsilon}\right) + V_2(x, y)\, f_2\left(z + t/\sqrt{\mu\varepsilon}\right) \tag{17.231}$$

(z being the space variable here)

where V_1 and V_2 are the solutions of the two-dimensional Laplace's Eq. (17.222a) and W_1 and W_2 are the complex potential functions related to V_1 and V_2 by Eqs. (17.223).

At this stage, we define a new two-dimensional vector operator ∇_2 which is

$$\nabla_2 = \mathbf{i}_x \frac{\partial}{\partial x} + \mathbf{i}_y \frac{\partial}{\partial y} \tag{17.232}$$

so that

$$\nabla_2 V = \mathbf{i}_x \frac{\partial V}{\partial x} + \mathbf{i}_y \frac{\partial V}{\partial y} \tag{17.233}$$

then

$$-\mathbf{i}_z \times \nabla_2 V = \nabla_2 U \qquad \text{and} \qquad \mathbf{i}_z \times \nabla_2 U = \nabla_2 V$$

and

$$\nabla \times \left[\nabla_2 U\{f(z)\}\right] = \mathbf{i}_z \times \nabla_2 \left[U\{f'(z)\}\right] \tag{17.234}$$

from Eq. (17.225).

From Eqs. (13.77) and (13.78), the vector potential for a transverse electric field can be written as

$$\begin{aligned}\mathbf{A} &= \nabla \times (\mathbf{i}_z W) \\ &= -\mathbf{i}_z \times \nabla_2 \, V_1(x, y) \, f_1\left(z - t/\sqrt{\mu\varepsilon}\right) - \mathbf{i}_z \times \nabla_2 V_2(x, y) \, f_2\left(z + t/\sqrt{\mu\varepsilon}\right) \\ &= \nabla_2 U_1(x, y) \, f_1\left(z - t/\sqrt{\mu\varepsilon}\right) + \nabla_2 U_2(x, y) \, f_2\left(z + t/\sqrt{\mu\varepsilon}\right)\end{aligned} \tag{17.235}$$

Obviously the first term above represents waves travelling in the +ve z-direction and the second term is for the waves travelling in the –ve z-direction.

From the vector potential, the magnetic and the electric fields would be derived as:

$$\mathbf{B} = \nabla \times \mathbf{A} = \nabla_2 V(x, y) f'\left(z \mp t/\sqrt{\mu\varepsilon}\right) \tag{17.236}$$

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} = \pm \frac{1}{\sqrt{\mu\varepsilon}} \nabla_2 U(x, y) f'\left(z \mp t/\sqrt{\mu\varepsilon}\right) \tag{17.237}$$

The +ve wave takes the upper sign in the above two equations. Combining Eqs. (17.234), (17.236) and (17.237), we get

$$\sqrt{\mu\varepsilon}\, \mathbf{E} = \mp \mathbf{i}_z \times \mathbf{B} \tag{17.238}$$

In a particular system, let it be assumed that the current goes out on one set of conducting cylinders and returns on another set. Equation (17.237) shows that all the members of one set would have the same potential at one value of z (the space variable) and hence for any other value of z. The relation between the total current and the charge (per unit length) in either set [of members, n in number (say)] can be found by using Gauss' theorem, as

$$Q = \sum_{i=1}^{n} \oint \varepsilon E_n \, ds_i = \sqrt{\mu\varepsilon} \sum_{i=1}^{n} \oint H_i ds_i = \sqrt{\mu\varepsilon}\, I \tag{17.239}$$

If L is the self-inductance per unit length, C the capacitance per unit length and S is the area outside the conductors in a plane of constant z, then

$$\frac{LI^2}{2} = \iint_S \frac{B^2}{2\mu} dS = \iint_S \frac{\varepsilon E^2}{2} dS = \frac{Q^2}{2C} = \frac{\mu\varepsilon}{2C} I^2 \tag{17.240}$$

Hence L and C are related as

$$LC = \mu\varepsilon = \frac{1}{\mathrm{v}^2} \tag{17.241}$$

Thus, LC is the reciprocal of the square of the electromagnetic wave velocity in the medium outside the conductors.

So far we have started from the vector potential and evaluated **E** and **B**. We now consider the inverse problem which is that **E** and **B** are specified either at a point in space as a function of time or as a function of space variable z at a given instant of time.

First we consider the case where **E** and **B** are specified in the plane $z = 0$ as functions of time. Then **A** for any z can be written down from Eq. (17.235). Thus,

$$\mathbf{A} = \frac{1}{2}\nabla_2 U(x, y)\left[f\left(t - z\sqrt{\mu\varepsilon}\right) + f\left(t + z\sqrt{\mu\varepsilon}\right) - g\left(t - z\sqrt{\mu\varepsilon}\right) - g\left(t + z\sqrt{\mu\varepsilon}\right)\right] \quad (17.242)$$

If U and V are given by Eqs. (17.223), then the fields at $z = 0$ can be obtained from **A** as:

$$\left.\begin{aligned} \mathbf{E}_{z=0} &= -\left(\frac{\partial \mathbf{A}}{\partial t}\right)_{z=0} = -\nabla_2 U(x, y)\, f'(t) \\ \mathbf{B}_{z=0} &= (\nabla \times \mathbf{A})_{z=0} = -\sqrt{\mu\varepsilon}\,\nabla_2\, V(x, y)\, g'(t) \end{aligned}\right\} \quad (17.243)$$

On the other hand, when **E** and **B** are given as functions of z at $t = 0$, **A** can again be written from Eq. (17.235) as:

$$\mathbf{A} = \frac{1}{2}\nabla_2 U\,(x, y)\left[f\left(z\sqrt{\mu\varepsilon} + t\right) - f\left(z\sqrt{\mu\varepsilon} - t\right) - g\left(z\sqrt{\mu\varepsilon} + t\right) - g\left(z\sqrt{\mu\varepsilon} - t\right)\right] \quad (17.244)$$

As before, if U and V are given by Eq. (17.223), then from **A** at $t = 0$, the fields can be expressed as

$$\mathbf{E}_{t=0} = -\nabla_2 U(x, y)\, f'\left(z\sqrt{\mu\varepsilon}\right), \quad \mathbf{B}_{t=0} = -\sqrt{\mu\varepsilon}\,\nabla_2 V(x, y)\, g'\left(z\sqrt{\mu\varepsilon}\right) \quad (17.245)$$

17.18.1 Propagation in a Medium of Finite Conductivity (Intrinsic Impedance)

So far the dielectric medium of propagation was assumed to be loss-less. Now we consider a medium of finite conductivity $(= \sigma)$ or finite resistivity $\left(= \rho = \frac{1}{\sigma}\right)$. In such a medium, the scalar wave propagation Eq. (17.221) modifies to

$$\left(\frac{\partial^2 W}{\partial x^2} + \frac{\partial^2 W}{\partial y^2}\right) + \left(\frac{\partial^2 W}{\partial z^2} - \mu\sigma\frac{\partial W}{\partial t} - \mu\varepsilon\frac{\partial^2 W}{\partial t^2}\right) = 0 \quad (17.246)$$

In this equation, when the second bracket is equated to zero, the solution is no longer a simple one-dimensional wave propagation due to the presence of the term $\mu\sigma\frac{\partial W}{\partial t}$. Writing its solution as

$$W = ZT, \quad \text{where } Z = Z(z), \quad T = T(t) \quad (17.247)$$

the equation becomes

$$\frac{Z''}{Z} = \mu\sigma\frac{T'}{T} + \mu\varepsilon\frac{T''}{T} = K \leftarrow \text{separation constant} \quad (17.248)$$

The solution is exponential in form and it can take either of the two forms, i.e.

(i) If the separation constant is real, the solution will be harmonic is space and will be a 'transient' state solution;
(ii) If the separation constant is imaginary, the solution will have harmonics in time and it will be a 'steady-state' behaviour solution.

We now consider the steady-state solutions in which the time-variation is given by $\exp(j\omega t)$. The function $f\left(z - t/\sqrt{\mu\varepsilon}\right)$ of Eq. (17.230) is now replaced by $\exp(\pm z\,\Gamma + j\omega t)$ where

$$\Gamma^2 = j\omega\mu\,(\sigma + j\omega\varepsilon) = (\alpha + j\beta)^2 \tag{17.249}$$

17.18.2 Solution of Propagation Equation in Cylindrical Coordinates

In Section 17.18, we solved for the special type of cylindrical waves moving in the z-direction obtained by setting the terms of the two parts of the scalar wave equation (17.221) to separately equal to zero, i.e.

$$\left(\frac{\partial^2 W}{\partial x^2} + \frac{\partial^2 W}{\partial y^2}\right) + \left(\frac{\partial^2 W}{\partial z^2} - \mu\varepsilon\frac{\partial^2 W}{\partial t^2}\right) = 0 \tag{17.221}$$

The above equation was reduced to two equations by equating each bracket (shown above) to zero separately, i.e.

$$\frac{\partial^2 W}{\partial x^2} + \frac{\partial^2 W}{\partial y^2} = 0 \qquad \text{and} \qquad \left(\frac{\partial^2 W}{\partial z^2} - \mu\varepsilon\frac{\partial^2 W}{\partial t^2}\right) = 0 \tag{17.222}$$

But this is not the only way to solve Eq. (17.221). On the other hand, the first equation of (17.222) can be equated to $\pm\beta_{mn}^2$, the second equation to $\mp\beta_{mn}^2$ and the time variation can be assumed to be of the sinusoidal variation type, i.e. $\exp(j\omega t)$ so that the second equation would then become

$$\frac{\partial^2 W}{\partial z^2} + \omega^2\mu\varepsilon W = \mp\beta_{mn}^2 \tag{17.259}$$

We can write $\omega^2\mu\varepsilon = \beta^2$ and scalar function W can be considered as $W = UZ$, where $U = U(x, y)$ and $Z = Z(z)$.

Then Eqs. (17.222) reduce to the form

$$\nabla_2^2 U \pm \beta_{mn}^2 U = 0 \quad \text{and} \quad \frac{d^2 Z}{dz^2} + (\beta^2 \mp \beta_{mn}^2)\,Z = 0 \tag{17.260}$$

where

$$\nabla_2^2 \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$$

If now instead of Cartesian coordinate system, a cylindrical coordinate system (r, ϕ, z) is used and the substitution

$$k_{mn}^2 = \beta^2 - \beta_{mn}^2 \quad \text{and} \quad k'^2_{mn} = \beta^2 + \beta_{mn}^2 \tag{17.261}$$

then the solution for W will be of the forms:

$$W = \{A\, e^{jk_{mn} z} + B\, e^{-jk_{mn} z}\}\, \{CJ_m(\beta_{mn} r) + DY_m(\beta_{mn} r)\} \cos(m\phi + \delta_m) \tag{17.262}$$

$$W = \{Ae^{jk'_{mn} z} + Be^{-jk'_{mn} z}\}\, \{CI_m(\beta_{mn} r) + DK_m(\beta_{mn} r)\} \cos(m\phi + \delta_m) \tag{17.263}$$

When C, D, k_{mn}, k'_{mn} are all real, both these equations represent waves propagating only in the z-direction. If $\beta_{mn}^2 > \beta^2$ so that k_{mn} is imaginary, Eq. (17.262) represents a wave exponentially damped in the z-direction. If C is real and D is complex, then Eq. (17.262) has a radial propagation component. If z is absent (i.e. no z variation) and $k_{mn} = 0$ and $k'_{mn} = 0$, then Eqs. (17.262) and (17.263) represent cylindrical wavefronts.

Defining transverse electric and transverse magnetic waves as those whose electric and magnetic fields, respectively, are normal to the direction of propagation (which is z-direction in the present case), then

$$\mathbf{A} = \frac{\mathbf{i}_r}{r}\frac{\partial W_{\text{te}}}{\partial \phi} - \mathbf{i}_\phi\, \phi \frac{\partial W_{\text{te}}}{\partial r} - \left[\mathbf{i}_r \frac{\partial^2 W_{\text{tm}}}{\partial r\, \partial z} + \mathbf{i}_\phi\, \phi \frac{\partial^2 W_{\text{tm}}}{r \partial \phi \partial z} + \mathbf{i}_z \left(\beta^2 W_{\text{tm}} + \frac{\partial^2 W_{\text{tm}}}{\partial z^2}\right)\right] \tag{17.264}$$

$$\mathbf{B} = \mathbf{i}_r \frac{\partial^2 W_{\text{te}}}{\partial r \partial z} + \mathbf{i}_\phi\, \phi \frac{\partial^2 W_{\text{te}}}{r \partial \phi \partial z} + \mathbf{i}_z \left(\beta^2 W_{\text{te}} + \frac{\partial^2 W_{\text{te}}}{\partial z^2}\right) + \mathbf{i}_r \frac{\beta^2}{r}\frac{\partial W_{\text{tm}}}{\partial \phi} + \mathbf{i}_\phi\, \phi\, \beta^2 \frac{\partial W_{\text{tm}}}{\partial r} \tag{17.265}$$

where
$$\left.\begin{aligned} \alpha &= \sqrt{\left[\left(\frac{\omega\mu}{2}\right)\left\{\sqrt{(\omega^2\varepsilon^2 + \sigma^2)} - \omega\varepsilon\right\}\right]} \\ \beta &= \sqrt{\left[\left(\frac{\omega\mu}{2}\right)\left\{\sqrt{(\omega^2\varepsilon^2 + \sigma^2)} + \omega\varepsilon\right\}\right]} \end{aligned}\right\} \tag{17.250}$$

Here Γ = the propagation constant

α = the attenuation constant

β = the phase constant or wave number.

The expression for **A** now becomes

$$\begin{aligned} \mathbf{A} &= \nabla \times \mathbf{i}_z W \\ &= \pm\, \mathbf{i}_z \times \nabla_2 V(x, y) \exp(\mp \Gamma z + j\omega t) \\ &= \mp\, \nabla_2 U(x, y) \exp(\mp \Gamma z + j\omega t) \end{aligned} \tag{17.251}$$

Hence the **B** and the **E** fields will be:

$$\begin{aligned} \mathbf{B} &= \nabla \times \mathbf{A} \\ &= \text{real part} \mp \Gamma \nabla_2 V(x, y) \exp(\mp \Gamma z + j\omega t) \\ &= \mp\, \nabla_2 V(x, y)\{\exp(\mp \alpha z)\}\, \{\alpha \cos(\omega t \mp \beta z) - \beta \sin(\omega t \mp \beta z)\} \end{aligned} \tag{17.252}$$

$$\begin{aligned} \mathbf{E} &= -\frac{\partial \mathbf{A}}{\partial t} \\ &= \text{real part} - j\omega \nabla_2 U(x, y) \exp(\mp \Gamma z + j\omega t) \end{aligned}$$

$$= \omega \nabla_2 U(x, y) \{\exp(\mp \alpha z)\} \{\sin(\omega t \mp \beta z)\} \tag{17.253}$$

and

$$\Gamma \mathbf{E} = \mp j\omega(\mathbf{i}_z \times \mathbf{B}) \tag{17.254}$$

or

$$\Gamma \mathbf{i}_z \times \mathbf{E} = \pm j\omega \mathbf{B}$$

From the above Eqs. (17.251), (17.252) and (17.253), it is obvious that these fields are not plane waves, but are attenuated (or damped) waves and their amplitudes are exponentially attenuated by the factor $e^{-\alpha z}$ as the waves progress in the z-direction.

If now, we define

R = the resistance per unit length between the conductors, and

C = the capacitance per unit length,

then
$$CR = \rho\varepsilon = \frac{\varepsilon}{\sigma} \tag{17.255}$$

(Ref: Problem 4.31 in *Electromagnetism: Problems with Solutions*, 2nd Edition, 2008)

The shunt admittance Y and the series impedance Z_L are given by

$$Y = \frac{1}{R} + j\omega C = \frac{\sigma + j\omega\varepsilon}{\varepsilon} \frac{\mu\varepsilon}{L} = \frac{\Gamma^2}{j\omega L}, \qquad \text{where } Z_L = j\omega L \tag{17.256}$$

Hence the characteristic impedance of the line will be

$$Z_c = \sqrt{\frac{Z_L}{Y}} = \frac{j\omega L}{\Gamma} = \frac{V}{I} = \frac{LE}{B} = \frac{j\omega\mu\varepsilon}{\Gamma C} \tag{17.257}$$

When the conductors consist of two infinite parallel planes 1 metre apart, then since an infinite tubular section of 1 m^2 running in the z-direction has C per unit length as equal to ε, the characteristic impedance will be

$$Z_c = \sqrt{\left(\frac{j\omega\mu}{s + j\omega\varepsilon}\right)} = \frac{j\omega\mu}{\alpha + j\beta} \tag{17.258}$$

PROBLEMS

17.1 For a uniform plane wave in air the magnetic field is given by

$$\mathbf{H} = \mathbf{i}_z \, 2 \exp\,[j\{\omega t - \pi z/(20)\}]$$

Calculate (i) the wavelength, (ii) the frequency, and (iii) the value of **E** at 1/(15) μsec, z = 5 m.

Ans.: 40 m, 7.5 MHz, $\mathbf{i}_x(-533)$ V/m

17.2 A 5 GHz plane wave is propagating in a large block of polystyrene (ε_r = 2.5), the amplitude of the electric field being 10 mV/m. Find

(i) the velocity of propagation,
(ii) the wavelength, and
(iii) the amplitude of the magnetic field intensity.

Ans.: 1.896×10^8 m/sec, 3.79 cm, 41.9 μA/m

17.3 The amplitude of the electric field component of a sinusoidal plane wave in free space is 20 V/m. Calculate the power per square metre carried by the wave.

Ans.: 0.53 W/m^2

17.4 A plane, linearly-polarized wave $\mathbf{E}_i$, $\mathbf{H}_i$, in free space, as described by the equations

$$\mathbf{E}_i = \mathbf{i}_x E_0 \exp\,[j(\omega t - \beta z)], \qquad \mathbf{H}_i = \mathbf{i}_y H_0 \exp\,[j(\omega t - \beta z)]$$

is incident on the plane surface ($z = 0$) of a semi-infinite block of loss-less dielectric of permittivity ε_r, and gives rise to a transmitted wave $\mathbf{E}_t$, $\mathbf{H}_t$, and a reflected wave $\mathbf{E}_r$, $\mathbf{H}_r$. This surface is coated with a thin layer of resistive material, of resistivity ρ_S, such that the thickness of this layer can be neglected. Show that the ratio of the amplitude of the reflected wave to that of the incident wave will be

$$= \frac{1 - Z_0\left(\dfrac{1}{\rho_S} + \dfrac{1}{Z_r}\right)}{1 + Z_0\left(\dfrac{1}{\rho_S} + \dfrac{1}{Z_r}\right)}$$

where $Z_0 = \sqrt{\dfrac{\mu_0}{\varepsilon_0}}$, $\qquad Z_r = \sqrt{\dfrac{\mu_0}{\varepsilon_r}}$.

What will be this ratio, if this layer of resistive material is removed from the incident surface?

17.5 A slab of solid dielectric material is coated on one side with a perfectly conducting sheet. A uniform, plane, sinusoidal wave is directed towards the uncoated side at normal incidence. Show that, if the frequency is such that the thickness of the slab is half a wavelength, the wave reflected from the dielectric surface will be equal in amplitude to the incident wave and opposite in phase. Calculate this frequency for a loss-less dielectric of permittivity 2.5 and thickness 5 cm.

Ans.: (i) Necessary condition is $k\beta d = \pi$, where $d = \left(\dfrac{1}{2}\right)$wavelength, and velocity in the dielectric medium (2) = c/k, $k = \sqrt{2.5}$

(ii) For $d = 5$ cm, $f = 1900$ MHz.

17.6 A plane wave of angular frequency ω in free space (μ_0, ε_0) is incident normally on a half-space of a very good conductor (μ_0, ε_0, σ). Show that the ratio of the reflected to the incident time-averaged Poynting vector is approximately

$$R_S = 1 - 2\beta\delta$$

where $\beta = \omega\sqrt{\mu_0\varepsilon_0}$ and $\delta = \sqrt{\dfrac{2}{\omega\mu\sigma}}$.

17.7 A time-harmonic, plane wave is incident normally on a planar resistive sheet which is the plane interface $z = 0$ separating the half-space $z < 0$ (medium 1) from another half-space $z > 0$ (medium 2). Let the media 1 and 2 be characterized by the constitute parameters μ_1, ε_1 and μ_2, ε_2 respectively. A thin planar layer of resistive material is sandwiched between the half-space. The thickness of the layer is assumed to be very small compared with a wavelength so that it can be approximated by a sheet of zero

thickness, and can be assumed to occupy the $z = 0$ plane. The surface current density $\mathbf{J}_S$ on the resistor sheet and the **E** field tangential to it are related as follows:

$$\mathbf{J}_S = \sigma_S \mathbf{E}_t = \frac{\mathbf{E}_t}{\rho_S}$$

where

σ_S = surface conductivity = $1/\rho_S$

ρ_S = surface resistivity.

Show that the ratios of the reflected **E** wave and the transmitted **E** wave to the incident wave are

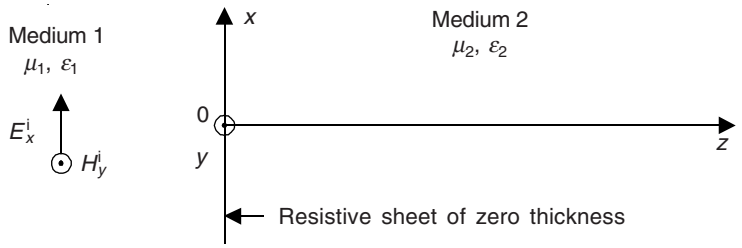
$$\rho_E = \frac{E^r_{xo}}{E^i_{xo}} = \frac{Z_{2r} - Z_1}{Z_{2r} + Z_1},$$

and

$$\tau_E = \frac{E^t_{xo}}{E^i_{xo}} = \frac{2Z_{2r}}{Z_{2r} + Z_1},$$

where

$$\frac{1}{Z_{2r}} = \frac{1}{\rho_S} + \frac{1}{Z_2}$$



Z_1, Z_2 = characteristic impedances of the media 1 and 2 respectively and the suffix o represents the amplitude of the corresponding wave. Hence show that the power dissipated in the resistive sheet per unit square area is

$$= \frac{\left|E^i_{xo}\right|^2}{2} \frac{\tau_E^2}{\rho_S}$$

The subscripts i and t have been made into superscripts here to eliminate confusion by overcrowding of suffices.

18 Waveguides

18.1 INTRODUCTION

In Chapter 17, we considered the propagation of electromagnetic waves, without referring to their sources or their destinations, and one of our conclusions was that they (the waves) are capable of transporting energy. In fact, the electromagnetic waves are used to transport energy from one place to another. An example of such application is the antenna of a television (or broadcasting) station, radiating the energy carrying information signals to all possible receivers in the vicinity. Naturally, a large portion of the energy of such waves is lost. Another and an important application is in the transmission of electromagnetic energy from a source to a single receiver. In such cases, it is desirable to minimize the transmission losses to the particular receiver. Since the electromagnetic waves propagating through free space are not the most suitable type for this purpose, it is necessary to design and devise some system to guide the energy of the wave along a desired route. All such systems can logically be called 'guides' of the electromagnetic waves, and the waves propagating along these guides are called the 'guided electromagnetic waves'. In Section 17.17.2, we were introduced to the basic concepts underlying the techniques of 'guiding' the electromagnetic waves.

Of the many such practical examples of 'guiding energy transfer', we shall quote only a few, starting with the first one which is also applicable at low (i.e. power frequency of 50 or 60 Hz) frequency. At these low frequencies, the electrical energy, which is the output of either a thermal or hydroelectric power station, is guided to various receivers by means of power transmission lines. Theoretically speaking, the power stations could radiate all that energy, and allow the consumers to hunt for as much of it, as per their requirement, from air. But such a method would be extremely uneconomical and possibly practically impossible.

At the higher end of the frequency spectrum, when a private message has to be sent from one person in one city to another person in another city (may even be in a different country), it would be very uneconomical and not the most efficient way of using a very powerful transmitter, creating electromagnetic waves all over the interlinked countries between the two cities. Instead the electromagnetic waves are propagated along the telephone lines, or optical fibres, and radio links with suitable signal amplifiers for regenerating the strength of the signals, thus compensating for the losses taking place during the transmission.

Another example is that of the transmission of extremely high (electromagnetic) pulses from a radar transmitter to its antenna. These energy pulses are sent through suitably designed conducting metal tubes (e.g. of rectangular or circular cross-sections), so that by guiding the waves to suitable paths, the energy losses are kept to a minimum. Such devices (i.e. these tubular pieces) are referred to as 'waveguides'. There are many varieties of waveguides for electromagnetic waves.

Though the types of electromagnetic waves propagating along the transmission lines and through the waveguides have some significant differences, they have some important features of similarity and there are underlying conceptual similarities in both.

In the transmission lines, the electromagnetic waves have both the **E** and the **H** vectors transverse to the direction of propagation, and are generally referred to as TEM (= Transverse Electromagnetic) waves. It should be noted that the coaxial cable problem discussed in Chapter 14, Section 14.4.2, was of the TEM wave type. However, in the transmission lines, **E** and **H** are not constants in a transverse plane, and hence the TEM waves along the transmission lines are *not* uniform. The most common types of transmission lines are the two-wire line, the coaxial cable, and the stripline.

The waveguides can be either metal tubes of rectangular or circular cross-section, or dielectric rods of different cross-sections. The electromagnetic waves propagating through them are not of the TEM type. Also, only the waves having a wavelength smaller than a certain critical wavelength, can propagate through the wave guides, the critical wavelength being a function of the transverse dimensions of the guides. The waveguides are used at very high frequencies.

18.2 TRANSMISSION LINES

The transmission lines, most often met with in practice, are made up of two (very nearly) parallel, cylindrical conductors. Since the conductor shapes can never be perfect, an exact rigorous analysis would be rather difficult. Hence, instead of treating the lines as a boundary value problem, the theory of travelling waves in lines (and also in cables) can be developed from the circuit standpoint. The lines are represented as a complicated combination of resistors, capacitors, and inductors. So, instead of the electric and the magnetic field intensities at all points where the fields of the lines exist, we consider the unknown voltage between the two lines and the current through them.

18.2.1 Transmission Lines with Negligible Losses (Loss-less Lines)

We consider the cable as a pair of conductors, in which each length δz has the inductance $L\delta z$, and between the conductors in each length δz is the capacitance $C\delta z$ (Figure 18.1). So, at a point A on the line (where its voltage is V),

$$\text{The charge in the capacitor is} = CV\delta z$$

And the current in the capacitor = the rate of increase of the charge

$$= C\left(\frac{\partial V}{\partial t}\right)\delta z$$

$\therefore$ Equating the inflowing and the outflowing currents at the point A (referring to Figure 18.1), we have

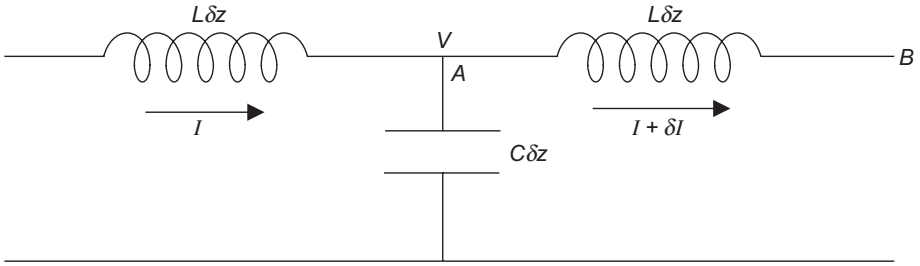


Figure 18.1 Circuit element representation of the transmission line.

$$I = C\left(\frac{\partial V}{\partial t}\right)\delta z + (I + \delta I)$$

or

$$\delta I + C\left(\frac{\partial V}{\partial t}\right)\delta z = 0$$

or, in the limit

$$\left(\frac{\partial I}{\partial z}\right) + C\left(\frac{\partial V}{\partial t}\right) = 0 \tag{18.1}$$

Next, considering the potentials along the line,

Potential at the point A − Potential at the point B $= L\delta z\left(\frac{\partial}{\partial t}\right)(I + \delta I) = -\delta V$

or

$$L\delta z\left(\frac{\partial I}{\partial t}\right) = -\delta V, \text{ neglecting higher order terms.}$$

In the limit,

$$\left(\frac{\partial V}{\partial z}\right) + L\left(\frac{\partial I}{\partial t}\right) = 0 \tag{18.2}$$

Eliminating I between Eqs. (18.1) and (18.2), we get

$$\left(\frac{\partial^2 V}{\partial z^2}\right) = \left(\frac{1}{\mathrm{u}^2}\right)\left(\frac{\partial^2 V}{\partial t^2}\right), \text{ where } \mathrm{u}^2 = \frac{1}{LC} \tag{18.3}$$

which is the usual one-dimensional wave equation.

It should be noted that L and C are related by the second equation of (18.3) and as the velocity of the wave u depends on the permittivity of the dielectric medium, the inductance and the capacitance are not so independent as the mode of deriving them would appear to be. From the earlier chapters, we have obtained that for the unit length of two parallel wires, their (L and C) expressions are

$$L = \left(\frac{\mu_0}{2\pi}\right)\ln\left(\frac{b}{a}\right) \quad \text{and} \quad C = \frac{-2\pi\varepsilon_0\varepsilon_\mathrm{r}}{\ln\left(\frac{b}{a}\right)}$$

neglecting the internal flux for the inductance.

The solution for the wave equation for the voltage V can be taken as

$$V = f(z - \mathrm{u}t)$$

and hence from Eq. (18.1),

$$I = \mathrm{u}Cf(z - \mathrm{u}t)$$

$\therefore$ We can write $I = \dfrac{V}{Z_c}$, (Z_c = characteristic impedance)

where

$$Z_c = \frac{1}{\mathrm{u}C} = \sqrt{\frac{L}{C}} \tag{18.4}$$

The value of the surge impedance Z for the air-cored cable is about 500 Ω and 30 Ω for a coaxial cable with solid insulation.

We shall now consider the behaviour of a line under two sets of different conditions:

1. *Transient shock condition.* This condition is obtained when a switch in the line is closed suddenly. Sudden switching and lightening are the two main sources, which cause the transient behaviour of the circuits. This is equivalent to the application of a unit step function to the line. This function is a wave which suddenly rises from $V = 0$ to $V = 1$ at the time $t = 0$, and then remains constant at that value thereafter (Figure 18.2). Any other types of shocks can be

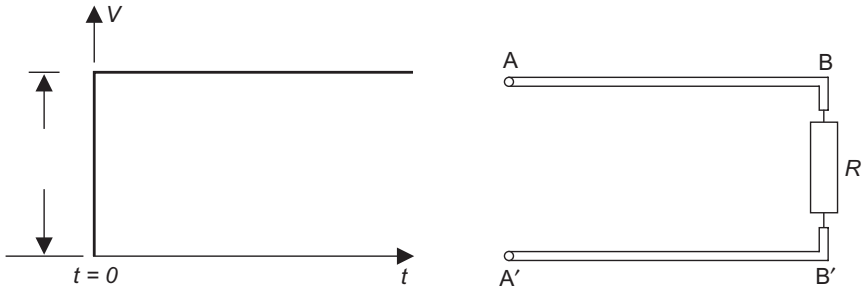


Figure 18.2 A unit-step function of voltage applied to the end AA′ of a transmission line whose other end BB¢ has been terminated by a resistance *R*.

built up by a successive superimposition of the unit steps. Let a voltage step of the type mentioned be applied to the terminals AA′ of a cable, whose other end BB′ is terminated by a resistance R as shown in Figure 18.2.

(1) The first effect of this step function is that a voltage wave of magnitude V and a current (V/Z_c) travel along the cable from A to B with a velocity u [Figure 18.3(a)].

(2) On reaching B, a reflected wave of voltage V' and a current (V'/Z_c) starts travelling from B to A. The magnitude of V' is determined by R such that (voltage/current) at BB′ = R. At this stage:

Potential drop across BB′ $= V + V'$, and

the current flowing out at $\mathrm{B} = \dfrac{V}{Z_c} - \dfrac{V'}{Z_c}$

$\therefore$ $$\frac{V + V'}{\dfrac{V - V'}{Z_c}} = R$$

and hence

$$V' = \left(\frac{R - Z_c}{R + Z_c}\right)V \qquad \text{[(Figure 18.3(b)]} \qquad (18.5)$$

There are three limiting cases of this stage, i.e.

(a) O.C. (open circuit) termination $R \to \infty$; $\therefore\ V' = V$ and $I' = I$
(b) S.C. (short circuit) termination $R = 0$; $\therefore\ V' = -V$ and $I' = I$
(c) Matched termination, $R = Z_c$; $\therefore\ V' = 0$ and $I' = 0$, i.e. no reflected wave.

(3) The reflected wave on reaching AA′, meets the generator impedance, which produces the unit function p.d.: this is ideally a zero impedance, but in fact negligibly small impedance. Hence this returning wave meets a S.C. condition at AA′. Thus for the next stage, a wave $-V'$ travels from A to B, as shown in Figure 18.3(c).

(4) On reaching B, it is reflected back according to (2), and the reflected wave is

$$= -\left(\frac{R - Z_c}{R + Z_c}\right)V' - \left(\frac{R - Z_c}{R + Z_c}\right)^2 V$$

as shown in Figure 18.3(d).

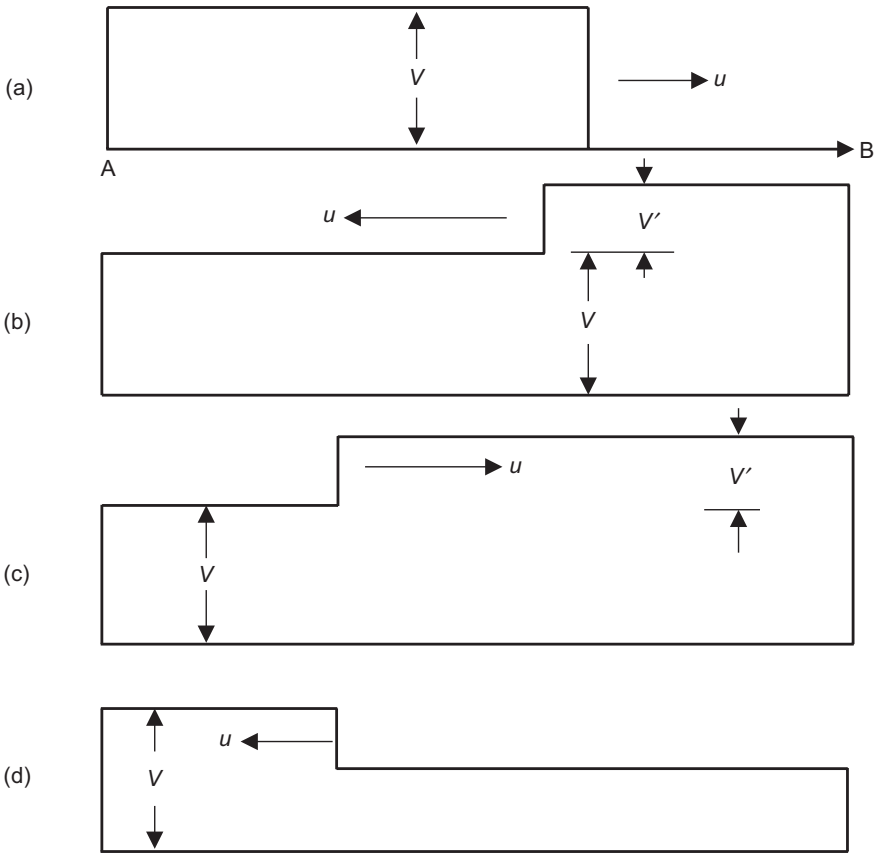


Figure 18.3 Repeated reflections on the transmission line.

With the passage of time, the cycle repeats the process. The voltage waveform at any point in the cable is thus obtained, starting from $t = 0$, and adding up the voltages associated with the waves that have arrived at that point, up to the instant of time under consideration. If T is the period of time required by the waves to traverse the length AB of the cable, then starting from $t = 0$ to $t = T$, there will be no voltage at these points on AB (Figure 18.4), i.e.

$V_i = 0$ from $t = 0$ to $t = T$

$V_i = V(1 + \alpha)$, where $\alpha = [(R - Z_c)/(R + Z_c)]$, from $t = T$ to $t = 3T$

$V_i = V(1 - \alpha^2)$, from $t = 3T$ to $t = 5T$

$V_i = V(1 + \alpha^3)$, from $t = 5T$ to $t = 7T$

$V_i = V(1 - \alpha^4)$, from $t = 7T$ to $t = 9T$

and so on. This is all shown in Figure 18.4.

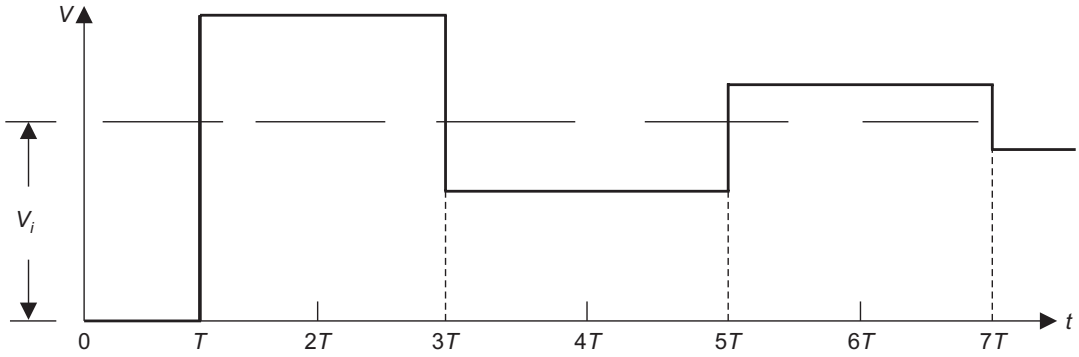


Figure 18.4 The voltage wave at any point on the cable AB.

2. *Steady-state A.C. operation.* In this condition, we replace the $(\partial/\partial t)$ operator in Eqs. (18.1) and (18.2) by the operator $(j\omega)$, so that these two equations become:

$$\frac{\partial I}{\partial z} + j\omega CV = 0 \quad \text{and} \quad \frac{\partial V}{\partial z} + j\omega LI = 0 \tag{18.6}$$

Combining these two equations,

$$\frac{\partial^2 V}{\partial z^2} + \omega^2 LCV = 0 \quad \text{or} \quad \frac{\partial^2 V}{\partial z^2} + \beta^2 V = 0 \tag{18.7a}$$

where

$$\beta^2 = \omega^2 LC = \frac{(2\pi f)^2}{u^2} \tag{18.7b}$$

(*Note:* if λ = wavelength of the line, then $f\lambda = u$ and $\beta = 2\pi/\lambda$.)

The solution of Eq. (18.7a) can be written as

$$V = A \exp(j\beta z) + B \exp(-j\beta z) \tag{18.8}$$

The factor $\exp(-j\beta z)$ represents a phase lag, which increases with z; and thus represents a wave travelling to $z \to +\infty$.

Similarly, exp ($j\beta z$) represents a wave travelling to $z \to -\infty$.

Since the modulus of exp ($\pm j\beta z$) is unity, it means that the amplitudes of these waves remain unchanged as they travel.

However the more convenient form of the solution of Eqs. (18.7) is

$$V = A' \cos(\beta z) + B' \sin(\beta z) \tag{18.9}$$

$\therefore$ From the second equation of (18.6),

$$I = -\left(\frac{1}{j\omega L}\right)\beta\,[-A' \sin(\beta z) + B' \cos(\beta z)]$$

$$= \left(\frac{j}{Z_c}\right)[-A' \sin(\beta z) + B' \cos(\beta z)] \tag{18.10}$$

For convenience, we rearrange the coordinate system, and place the origin at the receiving end, i.e. B is $z = 0$ and A is $z = -l$, where l = the length of the section AB of the cable. Further, we use the suffices S and R respectively to denote the quantities at the sending end ($z = -l$) and the receiving end ($z = 0$), i.e.

at the sending end ($z = -l$), the quantities are V_S, I_S; and

at the receiving end ($z = 0$), the quantities are V_R, I_R.

Evaluating the unknowns A' and B' from the conditions at $z = 0$, from Eqs. (18.9) and (18.10),

$$V_R = A' + 0 \qquad \text{and} \qquad I_R = \left(\frac{j}{Z_c}\right)(-0 + B')$$

$\therefore$ $$A' = V_R \qquad \text{and} \qquad B' = \frac{Z_c I_R}{j} = -jZ_c I_R$$

$\therefore$ $$V = V_R \cos(\beta z) - jZ_c I_R \sin(\beta z) \tag{18.11a}$$

and

$$I = I_R \cos(\beta z) - j\left(\frac{V_R}{Z_c}\right)\sin(\beta z) \tag{18.11b}$$

And, at the sending end ($z = -l$),

$$V_S = V_R \cos(\beta l) + jZ_c I_R \sin(\beta l) \tag{18.12a}$$

$$I_S = I_R \cos(\beta l) + j\left(\frac{V_R}{Z_c}\right)\sin(\beta l) \tag{18.12b}$$

We consider the case, when the receiving end is open-circuited. Then, $I_R = 0$.

$\therefore$ $$V = V_R \cos(\beta z) \qquad \text{and} \qquad I = -j\left(\frac{V_R}{Z_c}\right)\sin(\beta z)$$

This above shows the standing wave patterns for both V and I, and both of them oscillate in unison. The wave shapes are shown in Figure 18.5. The conditions at the sending end depend on the length of the line, and the wavelength of the wave. One case of interest is when $l = \lambda/4$, i.e. quarter wave line,

$$\beta l = \left(\frac{2\pi}{\lambda}\right)\left(\frac{\lambda}{4}\right) = \frac{\pi}{2}$$

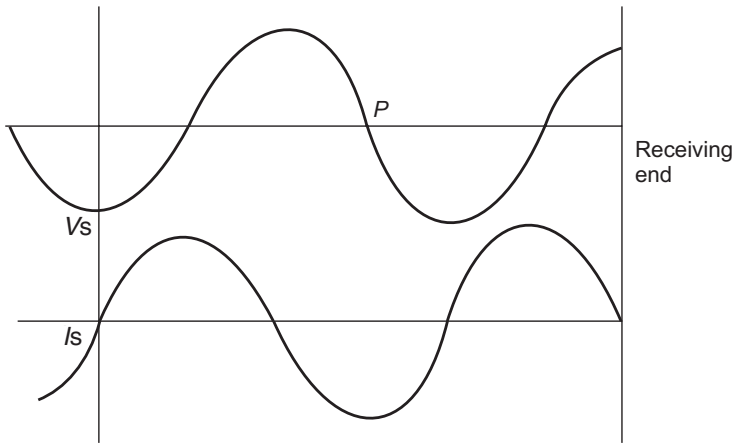


Figure 18.5 Standing wave patterns for V and I for an O.C. transmission line.

$\therefore$ From Eqs. (18.12),

$$V_S = jZ_cI_R, \qquad I_S = j\left(\frac{V_R}{Z_c}\right)$$

If we write $V_S = Z_SI_S$ and $V_R = Z_RI_R$, then, we get

$$Z_S = \frac{Z_c^2}{Z_R} \tag{18.13}$$

$\therefore$ If the receiving end is short-circuited, i.e. $Z_R = 0$, then $Z_S \to \infty$; i.e. the sending end appears to be open-circuited, and vice versa. Thus the quarter wavelength line can be used as an 'impedance transformer' to alter the apparent value of the impedance across a pair of terminals.

When a line is used for communication with high frequency currents, the standing waves of the type shown above are highly undesirable. This is because a detector (say, at a point P) will receive no signals. Such a state can be avoided, if the line is terminated with matched load, i.e. a pure resistance of magnitude equal to Z_c.

$\therefore$ $V_R = Z_cI_R$, and Eqs. (18.11) become

$$V = V_R[\cos(\beta z) - j\sin(\beta z)] = V_R \exp(-\beta z) \tag{18.14a}$$

$$I = I_R[\cos(\beta z) - j\sin(\beta z)] = I_R \exp(-\beta z) \tag{18.14b}$$

which are travelling waves [as $\exp(j\omega t)$ is implicit in these expressions] moving to $z \to +\infty$, and thus the reflected waves are eliminated.

Again, expressing the voltage and the current equations of (18.11) in terms of impedances, we have

$$V = I_R[Z_L\cos(\beta z) - jZ_c\sin(\beta z)] \tag{18.15a}$$

$$I = \left(\frac{I_R}{Z_c}\right)[Z_c\cos(\beta z) - jZ_L\sin(\beta z)] \tag{18.15b}$$

where the terminating resistance R has now been replaced by a more general load impedance Z_L, so that $Z_R = Z_L$.

Hence the impedance of the line (terminated by Z_L), at a distance z from the sending end, is

$$Z(z) = \frac{V(z)}{I(z)} = Z_c \left[\frac{Z_L - jZ_c \tan(\beta z)}{Z_c - jZ_L \tan(\beta z)} \right] \tag{18.16}$$

When the line is open-circuited, $Z_L \to \infty$; and

$$\therefore \quad Z(z) = \frac{jZ_c}{\tan(\beta z)} \tag{18.17}$$

and when the line is short-circuited, $Z_L = 0$; and

$$Z(z) = -jZ_c \tan(\beta z) \tag{18.18}$$

We notice that, in both the cases (i.e. O.C. and S.C.) the impedance of the line is purely imaginary, and hence they may replace any capacitor or inductor coil.

The reflection coefficient Γ is now given by

$$\Gamma = \frac{Z_L - Z_c}{Z_L + Z_c} \qquad \text{(the voltage reflection coefficient)*} \tag{18.19}$$

We can rewrite the voltage equation of (18.15) in terms of the reflection coefficient as

$$\begin{aligned} V(z) &= \left(\frac{1}{2}\right) I_R [(Z_L - Z_c) \exp(\beta z) + (Z_L + Z_c) \exp(-\beta z)] \\ &= \left(\frac{1}{2}\right) I_R [(Z_L + Z_c) \exp(\beta z) \{(\Gamma + \exp(-2\beta z)\}] \end{aligned} \tag{18.20}$$

Since $\left|\exp(-2\beta z)\right| = 1$, the ratio of the maximum to the minimum amplitude of the voltage along the line is

$$\frac{V_{max}}{V_{min}} = s = \frac{1 + |\Gamma|}{1 - |\Gamma|} \tag{18.21}$$

This ratio is known as the 'voltage standing wave ratio', which is abbreviated to VSWR. The above equation shows the range of s to be $1 \le s \le \infty$.

18.2.2 Lossy Transmission Lines

As before in the previous Section 18.2.1, we consider the transmission line cable as a pair of conductors of circular cross-section, which is characterized by the following parameters per unit length: resistance R (taking both the conductors into account), inductance L, capacitance C, and conductance G. The circuit representation of a section of this line is shown in Figure 18.6.

Considering the voltage and the current relations along the line, as shown in Figure 18.6 [similar to Eqs. (18.1) and (18.2) of the last section]:

$$V(z, t) - R\delta z I(z, t) - L\delta z \left(\frac{\partial}{\partial t}\right) I(z, t) = V(z, t) + \delta V(z, t)$$

and

$$I(z, t) - G\delta z\, [V(z, t) + \delta V(z, t)] - C\delta z \left(\frac{\partial}{\partial t}\right) [V(z, t) + \delta V(z, t)] = I(z, t) + \delta I(z, t)$$

*As distinct from the coefficients of energy reflection R_P and R_N of Chapter 17, p. 567.

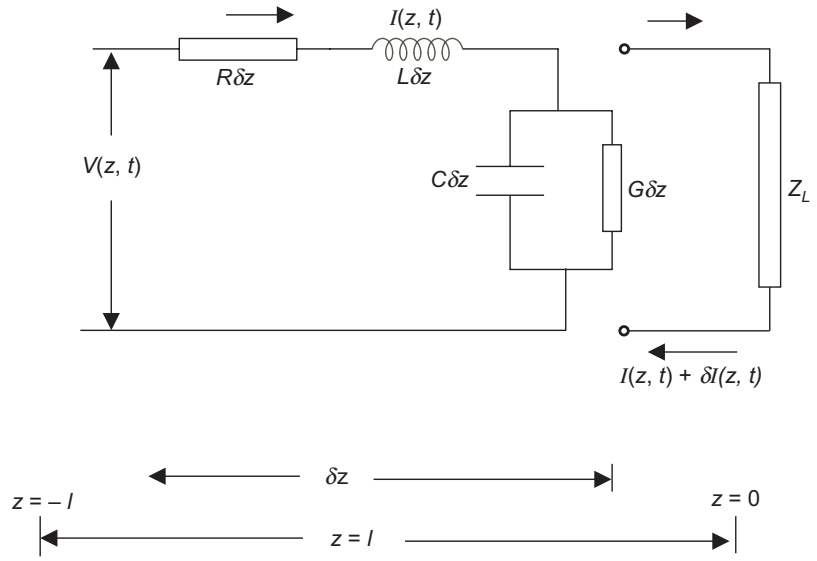


Figure 18.6 Circuit representation of a section of a lossy transmission line.

Simplifying and ignoring $(\delta z)^2$ and higher degree terms,

$$\left(\frac{\partial}{\partial z}\right)V(z, t) + RI(z, t) + L\left(\frac{\partial}{\partial t}\right)I(z, t) = 0 \tag{18.22a}$$

$$\left(\frac{\partial}{\partial z}\right)I(z, t) + GV(z, t) + C\left(\frac{\partial}{\partial t}\right)V(z, t) = 0 \tag{18.22b}$$

The most important case for practical problems is when $V(z, t)$ and $I(z, t)$ vary sinusoidally with time. So we can write:

$$I(z, t) = \text{Re}[I(z) \exp(j\omega t)] \qquad \text{and} \qquad V(z, t) = \text{Re}[V(z) \exp(j\omega t)] \tag{18.23}$$

where $I(z)$ and $V(z)$ are complex amplitudes of the current and the voltage along the line. Hence Eqs. (18.22) become:

$$\left(\frac{d}{dz}\right)V(z) + (R + j\omega L)I(z) = 0 \tag{18.24a}$$

$$\left(\frac{d}{dz}\right)I(z) + (G + j\omega C)V(z) = 0 \tag{18.24b}$$

Differentiating the first equation with respect to z and combining with the second; and doing the vice versa, we get

$$\left(\frac{d^2}{dz^2}\right)V(z) + (R + j\omega L)(G + j\omega C)V(z) = 0 \tag{18.25a}$$

$$\left(\frac{d^2}{dz^2}\right)I(z) + (R + j\omega L)(G + j\omega C)I(z) = 0 \tag{18.25b}$$

both of which are complex one-dimensional wave equations.

Let $$(R + j\omega L)(G + j\omega C) = \gamma^2 = (\alpha + j\beta)^2 \tag{18.26}$$

which gives

$$\alpha = \left(\frac{1}{\sqrt{2}}\right)\left[\sqrt{(R + j\omega L)(G + j\omega C)} + \left(RG - \omega^2 LC\right)\right]^{1/2} \tag{18.27a}$$

$$\beta = \left(\frac{1}{\sqrt{2}}\right)\left[\sqrt{(R + j\omega L)(G + j\omega C)} - \left(RG - \omega^2 LC\right)\right]^{1/2} \tag{18.27b}$$

where

γ is the complex propagation constant

α is the attenuation constant

β is the phase constant.

Equations (18.25) then become

$$\left(\frac{d^2}{dz^2}\right)V(z) + \gamma^2 V(z) = 0 \tag{18.28a}$$

$$\left(\frac{d^2}{dz^2}\right)I(z) + \gamma^2 I(z) = 0 \tag{18.28b}$$

The solution of both these equations is the same, i.e.

$$V(z) = V_1 \exp(-\gamma z) + V_2 \exp(\gamma z) \tag{18.29a}$$

$$I(z) = I_1 \exp(-\gamma z) + I_2 \exp(\gamma z) \tag{18.29b}$$

where V_1, V_2, I_1, and I_2 are unknown constants to be determined, but they are not all independent. These solutions appear as the sum of two waves, one travelling in the +ve z-direction and the other one in the –ve z-direction. Substituting from the above equations in Eqs. (18.24), we get

$$-\gamma V_1 \exp(-\gamma z) + \gamma V_2 \exp(\gamma z) + (R + j\omega L)[I_1 \exp(-\gamma z) + I_2 \exp(\gamma z)] = 0$$

This equation can be satisfied only if the coefficients of $\exp(-\gamma z)$ and $\exp(\gamma z)$ cancel out separately for all z, i.e.

$$\frac{V_1}{I_1} = \frac{R + j\omega L}{\gamma} = \sqrt{\frac{R + j\omega L}{G + j\omega C}} = Z_c \tag{18.30a}$$

$$\frac{V_2}{I_2} = -\frac{R + j\omega L}{\gamma} = -\sqrt{\frac{R + j\omega L}{G + j\omega C}} = -Z_c \tag{18.30b}$$

where Z_c is the characteristic impedance of the line.

If now we consider the transmission line to be terminated at the receiving end by a load impedance Z_L (= Z_R) which is located at $z = 0$ [i.e. the origin of the coordinate system being located at the receiving end where the load impedance is connected, and the –ve z-direction is towards the sending end, where the voltage source is connected (Figure 18.6), the length of this section of this line under consideration being l], then at $z = 0$,

$$Z_L\ (= Z_R) = \left[\frac{V(z)}{I(z)}\right]_{z=0} = \frac{V_1 + V_2}{I_1 + I_2}$$

$$= Z_c \left(\frac{I_1 - I_2}{I_1 + I_2}\right)$$

From this equation, the reflection coefficient ($= \Gamma_R$) is obtained as

$$\Gamma_R = \frac{V_2}{V_1} = \frac{Z_R - Z_c}{Z_R + Z_c} \tag{18.31a}$$

and

$$\frac{I_2}{I_1} = \frac{Z_c - Z_R}{Z_c + Z_R} \tag{18.31b}$$

Expressing the solutions of Eqs. (18.29) in hyperbolic function form, we have

$$V = A_1 \cosh(\gamma z) + B_1 \sinh(\gamma z) \tag{18.32a}$$

$$I = A_2 \cosh(\gamma z) + B_2 \sinh(\gamma z) \tag{18.32b}$$

where the unknowns A_1, B_1, A_2, B_2 are determined by the use of the boundary conditions at the receiving end ($z = 0$) and at the sending end ($z = -l$), as shown in Figure 18.7.

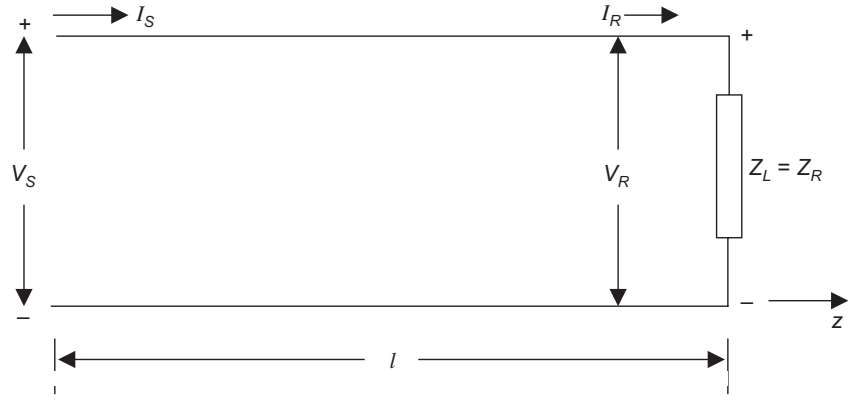


Figure 18.7 Transmission line terminated with a load impedance ($Z_L = Z_R$) at the receiving end.

The boundary conditions are:

(a) at $z = 0$, $V = V_R$, $I = I_R$; and

(b) at $z = -l$, $V = V_S$, $I = I_S$ (18.33)

Substituting these equations in (18.32), and also using Eqs. (18.24), we get

$$V_S = V_R \cosh(\gamma l) + Z_c I_R \sinh(\gamma l) \tag{18.34a}$$

$$I_S = I_R \cosh(\gamma l) + \left(\frac{V_R}{Z_c}\right) \sinh(\gamma l) \tag{18.34b}$$

These are general transmission line equations correlating the voltages and the currents at the two ends of the line. Hence the general expression for the input impedance of the line is

$$Z_{in} = \frac{V_S}{I_S}$$

$$= \frac{V_R \cosh(\gamma l) + Z_c I_R \sinh(\gamma l)}{I_R \cosh(\gamma l) + \left(\frac{V_R}{Z_c}\right)\sinh(\gamma l)} \tag{18.35}$$

From the general expression for the input impedance, we now look at certain special cases of interest, as in the previous section, i.e.

(1) a line, short-circuited at the receiving end, i.e. $Z_R = 0$

$\therefore$ $V_R = 0$, and the input impedance = $Z_{sc} = Z_c \tanh(\gamma l)$ (18.36)

(2) for an open-circuited line, $Z_R \to \infty$, $I_R = 0$

$\therefore$ the input impedance = $Z_{oc} = Z_c \coth(\gamma l)$ (18.37)

(3) the product of these two gives us: $Z_{sc}Z_{oc} = Z_c^2$ (18.38)

The low-loss or loss-less (idealized) line characteristics can also be obtained from the equations of this section, by considering the limit, when: $R << \omega L$ and $G << \omega C$.

Then, $Z = R + j\omega L = j\omega L$ and $Y = G + j\omega C = j\omega C$, and

$$Z_c = \sqrt{\frac{R + j\omega L}{G + j\omega C}} = \sqrt{\frac{L}{C}} \tag{18.39}$$

$$\gamma = \sqrt{(R + j\omega L)(G + j\omega C)} \simeq j\omega\sqrt{LC} = \alpha + j\beta$$

$\therefore$ $\alpha = 0$ and $\beta = \omega\sqrt{LC}$ (18.40)

For low-loss lines, the approximation for β is very good, but at times the approximation of zero for α may not be good enough, even though α is very small compared with β. A better approximation is obtained by rearranging the terms in γ, and using the binomial expansion as shown below:

$$\gamma = j\omega\sqrt{LC}\sqrt{\left(1 + \frac{R}{j\omega L}\right)\left(1 + \frac{G}{j\omega C}\right)}$$

$$= j\omega\sqrt{LC}\left(1 + \frac{R}{j2\omega L}\right) + \left(\frac{G}{j2\omega C}\right)$$

$$\simeq j\omega\sqrt{LC}\left(1 + \frac{R}{j2\omega L} + \frac{G}{j2\omega C}\right)$$

$$\simeq \frac{R}{2\sqrt{\frac{L}{C}}} + \frac{G\sqrt{\frac{L}{C}}}{2} + j\omega\sqrt{LC} \tag{18.41}$$

so that

$$\alpha \simeq \left(\frac{1}{2}\right)\left(\frac{R}{Z_c} + GZ_c\right) \quad \text{and} \quad \beta \simeq \omega\sqrt{LC} \tag{18.42}$$

This value of α needs to be used only for calculating the line losses. For the calculations of the voltage and the current distributions, the attenuation of most of the *low-loss ultra high frequency transmission lines* is so small that $\alpha = 0$ gives satisfactory results. For many purposes, the low-loss lines may be treated as loss-less, i.e. $R = G = \alpha = 0$. Hence the equations of the general transmission lines reduce to the form given below for the low-loss high frequency lines:

$$V_S = V_R \cos(\beta l) + jI_R Z_c \sin(\beta l) \tag{18.43a}$$

$$I_S = I_R \cos(\beta l) + j\left(\frac{V_R}{Z_c}\right)\sin(\beta l) \tag{18.43b}$$

where $Z_c = \sqrt{\frac{L}{C}}$ is a pure resistance.

The input impedance of such a line is

$$Z_S = \frac{V_S}{I_S} = Z_c\left[\frac{Z_R \cos(\beta l) + jZ_c \sin(\beta l)}{Z_c \cos(\beta l) + jZ_R \sin(\beta l)}\right] \tag{18.44}$$

18.2.3 Other Types of Transmission Lines

The coaxial cable has already been analyzed in detail in Chapter 14 (Sections 14.3.2, 14.4.1–14.4.2), and so we shall not repeat it here, even though it is a device to produce guided electromagnetic waves. However it is recommended that the students of the subject would find it beneficial to reread these above-mentioned sections, at this stage, following the knowledge gained so far, for further clarity of understanding.

The other commonly used transmission line is the 'stripline' which consists of two parallel strips of finite ε, μ, σ such that each strip's width b is much greater than its thickness a (i.e. $b >> a$). We shall not discuss this device in detail here.

Next we shall discuss some basic types of waveguides.

18.3 WAVEGUIDES

18.3.1 Introduction

So far, in this chapter, we have described some different types of transmission lines which have been defined as guiding systems having two parallel conductors (but there may be more conductors in a system as in three-phase power lines which have three or four parallel conductors). The currents in the two conductors at any cross-section are of equal magnitude, but opposite in direction, and hence the electric and the magnetic field vectors are in planes, which are transverse to the direction of the lines, and hence are also transverse to the direction of energy propagation. The waves propagating along the transmission lines are thus transverse electromagnetic waves (TEM).

Waveguides are guiding systems, which generally have the form of highly conducting tubes or dielectric rods of different cross-sections, i.e. rectangular, circular, etc., and the electromagnetic energy is transported through such systems. There are many types of waveguides. In this short introduction, we shall start with the type known as the rectangular waveguide, which has the form of a tube of rectangular cross-section.

Also, for the ease of analysis, we consider here an idealized system with perfectly conducting walls, and two-dimensional in nature, though in reality the waveguides are never perfectly straight nor infinitely long. However, in spite of all these limitations, the picture we derive on the basis of this idealized model, closely resembles the real situation.

The rectangular waveguides are metallic guides (as are the two-wire transmission lines), and these are used efficiently over frequencies ranging from few tens of kHz to few tens of GHz. As the frequencies go further towards optical spectrum, the metallic waveguides become inefficient because of excessive losses in them and then the dielectric waveguides are used in the form of optical fibres. Since these are cylindrical in structure, their behaviours are analysed by solving the operating equations in the cylindrical polar coordinate system. So the cylindrical waveguides are dielectric waveguides which are more useful in the sub-millimetre and optical wavelength range (i.e. in terms of frequency this will be in units of THz). However, there are two types of dielectric waveguides i.e. cylindrical waveguides (e.g. optical fibres) and slab waveguides which are of rectangular geometry and are used in thin films and integrated optical devices.

The waveguides and many derived structures are the basic building blocks of any microwave system. The engineering problems related to such structures are varied, and in our introductory discussion, we shall deal with only important basic concepts. An important concept which will be developed is the 'modal propagation' of waves in these guided structures.

18.3.1.1 Parallel plane waveguide

In Section 17.17.2, it was shown that when a uniform plane E.M. wave, with its E vector normal to the plane of incidence, hits a perfectly conducting plane obliquely, it produces a reflected wave which interacts with the incident wave to produce a wave pattern such that this wave pattern is stationary or standing in the direction normal to the reflecting surface (i.e. in this case in the z-direction as shown in Figures 17.23 and 17.24 of Section 17.17.2) and sliding parallel to the reflecting surface, i.e. in the x-direction.

The interference pattern (say, of **E** field) has zero value on the reflecting surface ($z = 0$ plane). There are successive parallel planes of zero E alternated by +ve and –ve crests of the **E**-wave as shown in Figure 17.24. It was further stated that if another conducting plane was introduced in this pattern along the plane of zero **E** [(say), the first plane of zeros from the $z = 0$], then in the gap between these two parallel conducting planes (of width $z = d$), there is an E-wave pattern which is stationary in the z-direction, and gliding parallel to the x-direction. So starting from a wave which hits the $z = 0$ surface at any oblique angle θ_i. a wave has been produced which travels in the x-direction, i.e. the wave has been guided along the x-direction. This is how the parallel plane waveguide has been developed. The mathematics of such a guided wave has been derived and discussed in Section 17.17.2 and so it will not be repeated here. Summarizing the results, the resultant electric and the magnetic field components are:

$$\left.\begin{aligned}\mathbf{E} &= -\mathbf{i}_y j2\, E_{0i+}\, \sin(\beta z\cos\theta_i)\exp[j(\omega t-\beta x\sin\theta_i)]\\ \mathbf{H} &= -\mathbf{i}_x j2\left(\frac{E_{0i+}}{Z_0}\right)\cos\theta_i\,\cos(\beta z\cos\theta_i)\exp[j(\omega t-\beta x\sin\theta_i)]\\ &\quad -\mathbf{i}_z 2\left(\frac{E_{0i+}}{Z_0}\right)\sin\theta_i\,\sin(\beta z\cos\theta_i)\exp[j(\omega t-\beta x\sin\theta_i)]\end{aligned}\right\} \tag{17.204}$$

These are travelling waves in the x-direction but standing (or stationary) patterns in the z-direction.

The electric field E_y which is parallel to the conducting boundary $z = 0$ (i.e. xy-coordinate plane), becomes zero on that plane, as well as parallel planes located at $z = d$ such that

$$\sin(\beta d\cos\theta_i) = 0 = \sin\pi \text{ or } \sin n\pi$$

$$\therefore \quad d = \frac{n\pi}{\beta\cos\theta_i}, \quad \beta = 2\pi f\sqrt{\mu_0\varepsilon_0}$$

$$\therefore \quad d = \frac{\pi}{\beta\cos\theta_i}$$

If λ is the wavelength of the incident wave (in the direction of its propagation), then $\lambda = 2\pi/\beta$ and hence

$$\left.\begin{aligned} d &= \frac{\pi}{\beta\cos\theta_i} = \frac{\lambda}{2\cos\theta_i}\\ \text{and more generally}\quad x = d &= \frac{n\pi}{\beta\cos\theta_i} = \frac{n\lambda}{2\cos\theta_i} \qquad n = 0, 1, 2, \ldots\end{aligned}\right\} \tag{17.205}$$

This means that the fields remain unaffected, if another conducting plane is interposed at any x given by Eq. (17.205). Hence we find that for a given angle of incidence θ_i, the height of the second conducting boundary, in order to maintain this described field pattern in the intermediate gap, will be a discrete multiple of $\lambda/(2\cos\theta_i)$. Or rewriting this equation by reversal,

$$\cos\theta_i = \frac{n\lambda}{2d} = \frac{n\pi}{\beta d} \tag{A}$$

So now we can say that when there are two parallel conducting boundaries separated by a distance d, the fields between them will be due to superposition of waves incident at angles θ_i given by the above equation. Since n is an integer, there will be discrete values of θ_i. It should be noted that when there is a single boundary, the field generated by the reflection of the incident wave would have survived at any angle of incidence, i.e. θ_i could change in a continuous manner. But when there are parallel boundaries, the boundary conditions on the two boundaries can be satisfied only by certain discrete values of θ_i and so the field pattern in the gap would survive only for these specific values of θ_i. So there have been some drastic changes in the process which are stated below:

(i) There has been a change from the continuous domain of θ to the discrete domain.

(ii) For a given boundary separation d and a specified frequency f, there are only a finite number of angles θ, as $\cos\theta \leq 1$.

(iii) When $d_n < \lambda/2$, then from Eq. (A) $\cos\theta$ becomes greater than 1 which is not possible. This means that no waves can be launched between the two conducting planes, if the gap between them is $< \lambda/2$, i.e. less than the half wavelength.

(iv) As d increases or λ decreases, the number of angles (i.e. θ) at which waves can be launched also increases.

In the field expressions [i.e. Eqs. (17.204)], $\cos\theta_i$ can be replaced by λ and dn from Eqs. (17.205), i.e.

$$\cos\theta_{\mathrm{i}} = \frac{n\lambda}{2d} = \frac{n\pi}{\beta d}$$

and

$$\beta z \cos\theta_{\mathrm{i}} = \frac{2\pi z}{\lambda}\frac{n\lambda}{2d} = \frac{n\pi z}{d}$$

and

$$\sin\theta_{\mathrm{i}} = \sqrt{\left[1 - \left(\frac{n\lambda}{2d}\right)^2\right]} = \frac{\beta_x}{\beta}$$

and the wave number (the phase constant) for the guided wave pattern $= \beta_x = \beta\sin\theta_{\mathrm{i}}$.

∴ The field existing between the two conducting boundaries can be written as

$$\left.\begin{aligned}
\mathbf{E} &= -\mathbf{i}_y\, j2\, E_{0i+}\sin\left(\frac{n\pi z}{d}\right)\exp[j(\omega t - \beta_x x)] \\
\mathbf{H} &= -\mathbf{i}_x\, 2\left(\frac{E_{0i+}}{Z_0}\right)\left(\frac{n\pi}{\beta d}\right)\cos\left(\frac{n\pi z}{d}\right)\exp[j(\omega t - \beta_x x)] \\
&\quad -\mathbf{i}_z\, j\,2\left(\frac{E_{0i+}}{Z_0}\right)\left(\frac{\beta_x}{\beta}\right)\sin\left(\frac{n\pi z}{d}\right)\exp[j(\omega t - \beta_x x)]
\end{aligned}\right\} \qquad (17.204a)$$

It should be noted that for a given value of n, the variation of the field in the z-direction is fixed and independent of the frequency. For example, for $n = 1$, the field pattern is half sine wave; and for $n = 2$, the field pattern is one complete cycle (both the bounding planes $z = 0$ and $z = d$, and so on.).

Thus the field pattern is unique for a given n, and that there is no gradual change from one pattern to another.

These unique field patterns are called the 'modal field patterns', and the propagation of these waves in the form of these discrete patterns is called 'modal propagation.'

Since the time-varying electric field cannot exist alone, it will be associated with orthogonal magnetic field. For $n = 0$, the E field will be zero, and so both H_x and H_z will also vanish.

The field components in this problem are such that the E field is transverse to the direction of propagation, and so it is called 'Transverse Electric' or TE mode. (m is put as a suffix to indicate the order of the mode and so n used earlier would be replaced by m.)

A similar 'Transverse Magnetic' or TM mode pattern would be obtained if the incident wave had the **E**-wave parallel to the plane of incidence which has been described in Section 17.17.1.

The corresponding field expressions for the TM mode would be

$$\left.\begin{aligned}\mathbf{E} &= -\mathbf{i}_x\, j2E_{0i+}\left(\frac{m\pi}{\beta d}\right)\sin\left(\frac{m\pi z}{d}\right)\exp[\,j(\omega t-\beta_x x)] \\ &\quad -\mathbf{i}_z\, 2E_{0i+}\left(\frac{\beta_x}{\beta}\right)\cos\left(\frac{m\pi z}{d}\right)\exp[\,j(\omega t-\beta_x x)] \\ \mathbf{H} &= \mathbf{i}_y\, 2\left(\frac{E_{0i+}}{Z_0}\right)\cos\left(\frac{m\pi z}{d}\right)\exp[\,j(\omega t-\beta_x x)]\end{aligned}\right\} \qquad (17.191')$$

by similar algebraic manipulations as for the TE mode expressions.

Note: The suffix *m* for either TE mode or TM mode (or TEM) denotes the order of the mode, i.e. the number of half-cycles of the wave pattern in the gap width *d*.

18.3.1.2 Cut-off frequency

The modal propagation constant or the wave number in the direction of propagation (i.e. *x*-direction) is

$$\beta_x = \beta\sin\theta_{\mathrm{i}} = \beta\sqrt{\left[1-\left(\frac{m\lambda_1}{2d}\right)^2\right]}$$

$$= \sqrt{\left[\beta^2-\left(\frac{m\pi}{d}\right)^2\right]}$$

From the field expressions (17.204a), it follows directly that if the resulting pattern is to be travelling in the *x*-direction [i.e. $\exp\{j(\omega t-\beta_x x)\}$] then β_x must be real. If β_x is not real, i.e. becomes imaginary, then $j\beta_x x$ becomes real and then the function does not represent a travelling wave. It then becomes an attenuated standing alternating pattern.

Hence the condition for wave propagation is

$$\beta_x = \text{Real} \Rightarrow \beta \geq \frac{m\pi}{d}$$

Since

$$\beta = \frac{2\pi}{\lambda} = \frac{2\pi f}{\mathrm{u}}$$

where u is the velocity of the uniform plane wave in the medium under consideration, and we get

$$f \geq \frac{m\mathrm{u}}{2d}$$

or

$$\lambda \leq \frac{2d}{m}$$

So for a given waveguide height d, the frequency f must be higher than a certain threshold value, for the wave to propagate for a particular mode. This threshold value of the frequency is called the cut-off frequency for that mode, i.e.

$$f_c (\text{or } f_{\text{cm}}) = \frac{m\text{u}}{2d_1}$$

and the corresponding cut-off wavelength $\lambda_c (\text{or } \lambda_{\text{cm}}) = \frac{2d}{m}$.

18.3.2 Rectangular Waveguides

The present analysis is essentially a continuation of the matter discussed in Section 17.17.2 (and should be read as such, since the present discussion is a direct follow up of that section), where we analyzed in detail the wave pattern produced by the oblique incidence of a plane wave, with its **E** vector normal to the plane of incidence, on a perfectly conducting surface. We found that parallel to the reflecting interface plane $z = 0$, there are planes like AA′ (referring to Figure 17.24) on which $E_{\text{i}} + E_{\text{r}} = 0$ for all x and y at all t. This means that the wave pattern can still exist, if one of these parallel planes is replaced by a second conducting wall (similar to the $z = 0$ plane). If the nearest of such planes to $z = 0$ is taken, i.e. A_1A_1' (Figure 17.24), then its distance d_1 from the reflecting surface is related with the free space wavelength (denoted by λ_0), i.e.

$$d_1 = \frac{\lambda_0}{2 \cos \theta_{\text{i}}} = \frac{\lambda_0}{2} \sec \theta_{\text{i}} \tag{18.45}$$

The above equation refers to Eq. (17.206) of Section 17.17.2 where λ the wavelength of the incident wave has now been expressed as λ_0, as the wave is in free space, and hence,

$$\lambda_0 = \frac{1}{\left(\sqrt{\mu_0 \varepsilon_0}\right) f} = \frac{\text{c}}{f}$$

Or, conversely, we can argue that, given two perfectly conducting planes d_1 distance apart, a wave pattern can be set up at an angular frequency ω, such that:

$$\omega = 2\pi f = \frac{2\pi \text{c}}{\lambda_0} = \frac{\pi \text{c}}{d_1 \cos \theta_{\text{i}}} \tag{18.46}$$

The field $E_{\text{i}} + E_{\text{r}}$ does not vary with y, and hence can terminate on the ± charges on the planes at $y = 0$ and $y = b$. Thus a 'rectangular waveguide' is produced.

The field patterns are as shown in Figures 18.8(a) and (b). These are shown in two orthogonal planes, i.e. y-z plane and x-y plane. The separate fields E_{i}, E_{r}, H_{i}, H_{r} and $\angle\theta_{\text{i}}$, which were used for generating this final pattern, are really parts of the building scaffolding, and are now omitted.

From Figure 18.8(b), it will be seen that the field alternates longitudinally (in the x-direction—the direction of propagation of the guided wave—as distinct from the direction of propagation of the incident wave). It will also be noted that the **E** field (the resultant field) is transverse and **H** has both the transverse and the longitudinal components. The field travels with a velocity v_x given by

$$\text{v}_x = \text{c cosec } \theta_{\text{i}} \tag{18.47}$$

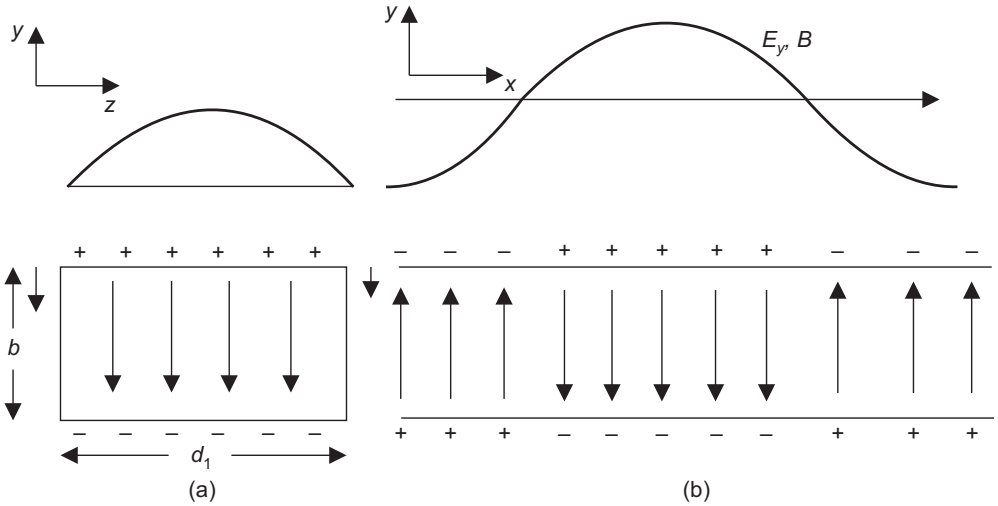


Figure 18.8 Field pattern in a rectangular waveguide (refer to Figure 17.24, for correspondence with the incident wave direction).

and the wavelength of the longitudinal waves is

$$\lambda_x = \lambda_0 \operatorname{cosec}\ \theta_i \tag{18.48}$$

(It should be noted that v_x is greater than c, and so v_x will exceed the velocity of light in an air-filled waveguide). It should be carefully noted that the velocity with which the signals are transmitted is the so called 'group velocity' v_{gr}, which is given by

$$v_{gr} = c \sin\ \theta_i \tag{18.49}$$

This is most simply explained by thinking of the transmission as by a sequence of reflected waves on two parallel planes $z = 0$ and $z = d_1$, as shown in Figure 18.9.

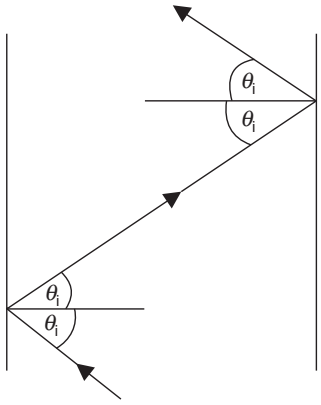


Figure 18.9 Group velocity v_{gr}.

To obtain the cut-off frequency, using Eqs. (18.45) and (18.48), we get

$$\left(\frac{\lambda_0}{2a}\right)^2 + \left(\frac{\lambda_0}{\lambda_x}\right)^2 = 1$$

or

$$\left(\frac{1}{\lambda_x}\right)^2 + \left(\frac{1}{2a}\right)^2 = \left(\frac{1}{\lambda_0}\right)^2 = \left(\frac{\omega}{2\pi c}\right)^2 \tag{18.50}$$

(We have replaced d_1 by a for simplicity of notation.)

$\therefore$ For λ_x to be real, $\frac{\omega}{\pi c}$ must be $> \frac{1}{a}$,

i.e. ω must be $> \frac{\pi c}{a}$.

$\therefore$ The limiting value of the cut-off frequency is given by

$$\omega_c = \frac{\pi c}{a} \tag{18.51}$$

which from Eqs. (18.46) and (18.49) corresponds to $\theta_i = 0$ and $v_{gr} = 0$.

Below this frequency there can be no propagation. This quantity relates the width of the waveguide to the proposed frequency.

If now we look at the lines of **E** and **H** in a plane containing the direction of propagation (as shown in Figure 18.10), then the lines of **E** (= E_y) in the x-y plane are lines transverse to the direction of propagation, and the lines of **H** are closed loops in the z-x plane (y = constant).

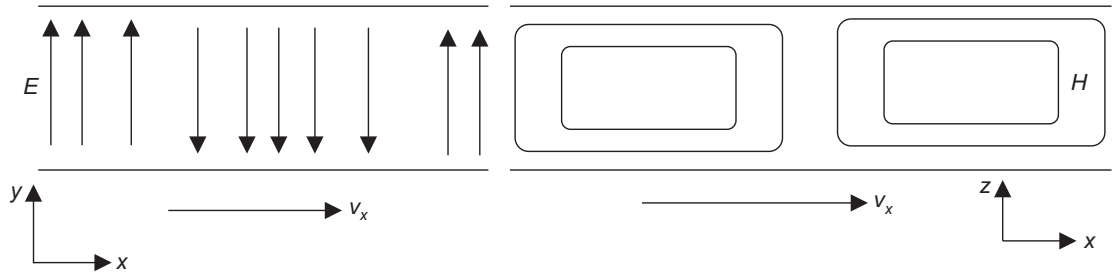


Figure 18.10 **E** and **H** lines along the direction of propagation: (a) **E** lines and (b) **H** lines.

The pattern, which we have discussed above, is only one of the many possible *modes of operation*, known as the TE_{10} mode where TE stands for the 'Transverse Electric', because the **E** field is entirely across the guide (i.e. totally normal to the direction of propagation). The suffix 10 is given because the standing wave of **E** has one-half cycle in the x-direction and none in the y-direction.

In general, there is a possibility of m, n, half-cycles in these two directions, leading to TE_{mn} mode. If we had chosen the second plane AA' (of Figure 17.24), then we should have obtained TE_{20} mode, with two half-cycles of **E** in the standing waves.

In addition, there is a whole series of modes in which the magnetic field is transverse; the TM_{11} being the lowest. The TM_{11} modes are as shown below in Figure 18.11.

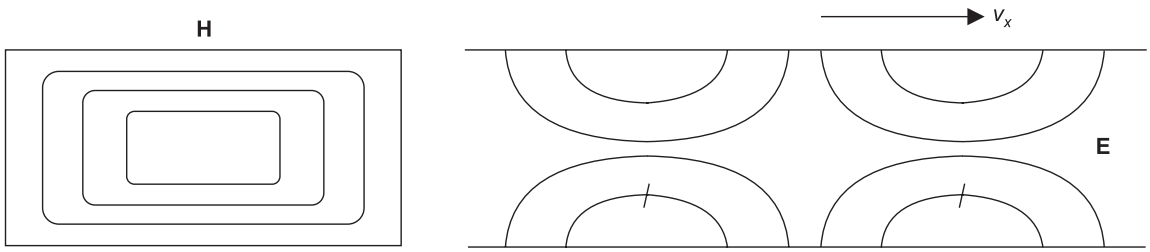


Figure 18.11 TM_{11} mode in a rectangular waveguide.

These broad types of waves can exist in all transmission systems. In a coaxial cable, for example the mode usually considered is TEM mode, but it can be shown that this cannot occur in a single-conductor waveguide (which is simply-connected). This is because a transverse magnetic field must have the basic form shown above (possibly with several vortices); $\oint \mathbf{H} \cdot d\mathbf{l}$ has a finite value round any closed line, showing that the current must be enclosed. This requires:

either (a) one or more central conductors (i.e. multiply-connected);

or (b) a displacement current, which necessitates a longitudinal electric field.

Thus TEM waves can only occur if there are conductors in the waveguides.

18.3.3 Rectangular Waveguide: Mathematical Analysis from the Fundamentals

In Section 18.3.2, we derived the behaviour of rectangular waveguides by starting from the reflection patterns of a plane wave incident obliquely on a perfect conducting surface. However, it is also possible to arrive at the same results directly, without using the pattern due to the oblique incidence of plane waves, and solve directly the Maxwell's equations for the specified geometry of the rectangular waveguide. We shall now analyze the device this way and confirm the results of the previous section. The advantage of such an approach is that, since the method of analysis is completely general, we can then suitably modify the results to account for the imperfections and changes in a practical device.

We start with the rectangular tube as shown in Figure 18.12, with its coordinate system shown in the figure, and the direction of propagation in the guide being $+z$-direction. The guide is

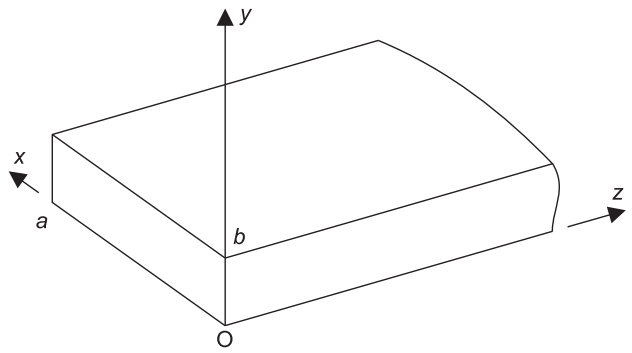


Figure 18.12 A rectangular waveguide, with cross-section $a * b$, and the coordinate system as shown.

assumed to be infinitely long, and its walls are (initially) made of perfect conductor, and the dielectric inside is loss-less; it has the characteristic properties: ε, and μ, with σ being zero.

It has been seen earlier that using complex notation for the sinusoidally time-varying excitations, a wave travelling in the $+z$-direction is denoted by $\exp(-j\beta z)$. When there is attenuation, $j\beta$ has to be replaced by γ, such that $\gamma = \alpha + j\beta$. So to maintain the generality, we indicate the propagation in the $+z$-direction by $\exp(-\gamma z)$. Hence the complex field vectors are of the form:

$$\mathbf{E} = \mathbf{E}(x, y) \exp(-\gamma z) \qquad \text{and} \qquad \mathbf{H} = \mathbf{H}(x, y) \exp(-\gamma z) \tag{18.52}$$

From the Maxwell's equations,

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} \qquad \text{and} \qquad \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \text{ we get}$$

$$\mathbf{i}_x\left(\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z}\right) + \mathbf{i}_y\left(\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x}\right) + \mathbf{i}_z\left(\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y}\right)$$
$$= j\omega\varepsilon\,(\mathbf{i}_x E_x + \mathbf{i}_y E_y + \mathbf{i}_z E_z)$$

or

$$\mathbf{i}_x\left(\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z}\right) + \mathbf{i}_y\left(\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x}\right) + \mathbf{i}_z\left(\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y}\right)$$
$$= -j\omega\mu\,(\mathbf{i}_x H_x + \mathbf{i}_y H_y + \mathbf{i}_z H_z)$$

using $\mathbf{B} = \mu\mathbf{H}$, $\mathbf{D} = \varepsilon\mathbf{E}$, and $\frac{\partial}{\partial t} = j\omega$.

Combining these equations with Eqs. (18.52), and equating the vector components, we get the following scalar equations:

$$\frac{\partial H_z}{\partial y} + \gamma H_y = j\omega\varepsilon E_x \tag{18.53a}$$

$$\frac{\partial H_z}{\partial x} + \gamma H_x = -j\omega\varepsilon E_y \tag{18.53b}$$

$$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = j\omega\varepsilon E_z \tag{18.53c}$$

$$\frac{\partial E_z}{\partial y} + \gamma E_y = -j\omega\mu H_x \tag{18.53d}$$

$$\frac{\partial E_z}{\partial x} + \gamma E_x = j\omega\mu H_y \tag{18.53e}$$

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = -j\omega\mu H_z \tag{18.53f}$$

From Eqs. (18.53), by combining (a) and (e) we get E_x and H_y, and by combining (b) and (d) we get H_x and E_y as given below:

$$H_x = -\left(\frac{\gamma}{k^2}\right)\left(\frac{\partial H_z}{\partial x}\right) + \left(\frac{\omega\varepsilon}{k^2}\right)\left(\frac{\partial E_z}{\partial y}\right) \tag{18.54a}$$

$$H_y = -\left(\frac{\gamma}{k^2}\right)\left(\frac{\partial H_z}{\partial y}\right) - \left(\frac{\omega\varepsilon}{k^2}\right)\left(\frac{\partial E_z}{\partial x}\right) \tag{18.54b}$$

$$E_x = -\left(\frac{\gamma}{k^2}\right)\left(\frac{\partial E_z}{\partial x}\right) - \left(\frac{\omega\mu}{k^2}\right)\left(\frac{\partial H_z}{\partial y}\right) \tag{18.54c}$$

$$E_y = -\left(\frac{\gamma}{k^2}\right)\left(\frac{\partial E_z}{\partial y}\right) + \left(\frac{\omega\mu}{k^2}\right)\left(\frac{\partial H_z}{\partial x}\right) \tag{18.54d}$$

where $k^2 = \gamma^2 + \omega^2\mu\varepsilon$.

The wave equations for E_z and H_z are obtained from the equations of (18.53) along with the Maxwell's divergence equations in charge-free region, i.e.

$$\nabla \cdot \mathbf{D} = 0 \qquad \text{and} \qquad \nabla \cdot \mathbf{B} = 0 \tag{18.55}$$

We differentiate (18.53a) with respect to y, (18.53b) with respect to x, and then add them and substitute from (18.53f) and (18.55) to get

$$\frac{\partial^2 H_z}{\partial x^2} + \frac{\partial^2 H_z}{\partial y^2} + \gamma^2 H_z = -\omega^2\mu\varepsilon H_z \tag{18.56}$$

and similarly differentiate (18.53d) with respect to y, (18.53e) with respect to x, then add them and substitute from (18.53e) and (18.55) to obtain:

$$\frac{\partial^2 E_z}{\partial x^2} + \frac{\partial^2 E_z}{\partial y^2} + \gamma^2 E_z = -\omega^2\mu\varepsilon E_z \tag{18.57}$$

Equations (18.54), (18.56) and (18.57) give us the relationship between the field components within the rectangular guide. It should be noticed that if both E_z and H_z are zero simultaneously, then all the fields within the guide will also vanish. Hence for such a waveguide structure (in which there is no inner conductor in the space of the guide—mathematically this region is 'singly-connected' whereas a region with conductor inside is 'multiply-connected'), there must exist either an E_z or a H_z component, i.e. *a* component directed in the direction of propagation. Thus two possible field configurations can exist in this waveguide, i.e. (1) the transverse electric (TE mode) waves for which $E_z = 0$, and (2) the transverse magnetic waves (TM mode) for which $H_z = 0$.

For the guide shown in Figure 18.12, the resultant boundary conditions are (we have assumed the enclosing boundaries to be perfect conductors, at this stage)

(1) at $y = 0$ and $y = b$, $E_x = E_z = 0$; and

(2) at $x = 0$ and $x = a$, $E_y = E_z = 0$.

Physically the conditions state that there can be no tangential electric field on any of the boundary surfaces.

18.3.3.1 Transverse electric waves in rectangular waveguides

In this type of waveguide, if $E_z = 0$, the electric field intensity vector is in the transverse plane, and such waves are referred to as transverse electric or TE waves.

When $E_z = 0$, Eqs. (18.53) simplify to:

$$\gamma E_y = -j\omega\mu H_x \tag{18.58d}$$

$$\gamma E_x = j\omega\mu H_y \tag{18.58e}$$

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = -j\omega\mu H_z \tag{18.58f}$$

$$\frac{\partial H_z}{\partial y} + \gamma H_y = j\omega\varepsilon E_x \tag{18.58a}$$

$$\frac{\partial H_z}{\partial x} + \gamma H_x = -j\omega\varepsilon E_y \tag{18.58b}$$

$$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = 0 \tag{18.58c}$$

We now express all the field components in terms of H_z, i.e. we combine Eqs. (18.58b) and (18.58d), and get

$$\frac{\partial H_z}{\partial x} + \gamma H_x = -j\omega\varepsilon\left[-j\left(\frac{\omega\mu}{\gamma}\right)H_x\right]$$

from which

$$H_x = -\left(\frac{\gamma}{k^2}\right)\left(\frac{\partial H_z}{\partial x}\right) \tag{18.59}$$

Similarly from (18.58a) and (18.58e), we get

$$H_y = -\left(\frac{\gamma}{k^2}\right)\left(\frac{\partial H_z}{\partial y}\right) \tag{18.60}$$

Combining these two equations with (18.58e) and (18.58d), we get

$$E_x = -\left(\frac{\omega\mu}{k^2}\right)\left(\frac{\partial H_z}{\partial y}\right) \tag{18.61}$$

$$E_y = \left(\frac{\omega\mu}{k^2}\right)\left(\frac{\partial H_z}{\partial x}\right) \tag{18.62}$$

and the equation for H_z is

$$\frac{\partial^2 H_z}{\partial x^2} + \frac{\partial^2 H_z}{\partial y^2} + k^2 H_z = 0 \tag{18.63}$$

We solve this equation by using the method of separation of variables, i.e.

$$H_z = H_z(x, y) = XY = X(x)Y(y) \tag{18.64}$$

Equation (18.63) becomes

$$\left(\frac{1}{X}\right)\left(\frac{\partial^2 X}{\partial x^2}\right) + \left(\frac{1}{Y}\right)\left(\frac{\partial^2 Y}{\partial y^2}\right) + k^2 = 0 \tag{18.65}$$

Since the first term is a function of x only, and the second term is a function of y only, the above equation can be satisfied only if each term is a constant, such that

$$\frac{\partial^2 X}{\partial x^2} = -k_x^2 X, \qquad \frac{\partial^2 Y}{\partial y^2} = -k_y^2 Y \tag{18.66}$$

where

$$k_x^2 + k_y^2 = k^2 = \gamma^2 + \omega^2 \mu\varepsilon \tag{18.67}$$

The solutions of Eqs. (18.66) are of the form

$$X = A_x \sin (k_x x) + B_x \cos (k_x x)$$

and

$$Y = A_y \sin (k_y y) + B_y \cos (k_y y)$$

so that

$$H_z = \sum_{k_x} \sum_{k_y} [A_x \sin (k_x x) + B_x \cos (k_x x)][A_y \sin (k_y y) + B_y \cos (k_y y)] \tag{18.68}$$

The unknown constants A_x, B_x, A_y, B_y, and k_x, k_y have to be determined by using the boundary conditions stated in the previous section.

The first set of boundary conditions state that

$$\text{at } y = 0, \text{ and } y = b, \; E_x = 0 \text{ and } E_z = 0$$

From Eqs. (18.61) and (18.68), we get

$$\sum_{k_x} \sum_{k_y} [A_x \sin (k_x x) + B_x \cos (k_x x)][A_y \cos (k_y y) - B_y \sin (k_y y)] k_y = 0$$

for $y = 0$ and $y = b$, for all values of x.

$\therefore \quad A_y = 0 \qquad \text{and} \qquad k_y = \dfrac{n\pi}{b}$, where $n = 0, 1, 2, 3, \ldots$.

The second set of boundary conditions states that at $x = 0$ and at $x = a$, $E_y = 0$ and $E_z = 0$.

$\therefore$ From Eqs. (18.62) and (18.68),

$$\sum_{k_x} \sum_{k_y} [A_x \cos (k_x x) - B_x \sin (k_x x)] k_x [A_y \sin (k_y y) + B_y \cos (k_y y)] = 0$$

for $x = 0$ and $x = a$ for all values of y.

$\therefore \quad A_x = 0$ and $k_x = \dfrac{m\pi}{a}$, where $m = 0, 1, 2, 3, \ldots$.

$$H_z = H_z(x, y) = H_0 \cos\left(\frac{m\pi x}{a}\right) \cos\left(\frac{n\pi y}{b}\right) \tag{18.69}$$

where m, $n = 0, 1, 2, 3, \ldots$, and H_0 stands for the product $(B_x B_y)$ of the constants from Eq. (18.68), and is determined by the intensity of the wave propagating through the guide. It should be noted that these boundary conditions cannot be satisfied if $H_z = 0$ all over the region inside the guide, and, if $H_z = 0$, then the equations define a wave for which **E** and **H** are constant in the transverse planes. So we have here proved mathematically that a rectangular waveguide cannot support a TEM wave.

All other components of TE type of waves, which can be supported by the waveguide, can now be calculated by using the equations from (18.59) to (18.62). An infinite number of modes corresponding to different values of m and n can propagate through the waveguide. The general

(mn) wave (or TE_{mn} mode, where the physical significance of m and n has been explained in Section 18.3.2) is supported by the waveguide, and the particular types are TE_{10}, TE_{01}, TE_{11}, TE_{12}, TE_{45}, etc. It should also be noted that the mode TE_{00} cannot exist, because then $H_z = H_0 =$ constant, so that then all the field components become zero.

From Eq. (18.67), we have

$$k^2 = k_x^2 + k_y^2 = \left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 = \gamma^2 + \omega^2\mu\varepsilon$$

$$\therefore \qquad \gamma = \left[\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 - \omega^2\mu\varepsilon\right]^{1/2} \tag{18.70}$$

In the transmission line theory, which we studied in the earlier part of this chapter, γ the propagation constant was found to be a complex number, i.e. $\gamma = \alpha + j\beta$, where α is the attenuation constant (or factor) (i.e. attenuation per unit length), and β is the phase-shift constant (i.e. phase-shift per unit length). If γ is real, then $\beta = 0$, i.e. there is no phase-shift along the tube, which means that for real γ [i.e. at low frequencies, since $\omega^2\mu\varepsilon$ is $< (m\pi/a)^2 + (n\pi/b)^2$] there is no wave motion along the guide. However as the frequency is increased, there is a critical value of ω ($= \omega_c$) at which the expression under the square-root sign is zero, and for higher values of ω, γ will be imaginary, i.e.

$$\gamma = j\beta = j\left[\omega^2\mu\varepsilon - \left\{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2\right\}\right]^{1/2}, \text{ when } \omega > \omega_c \tag{18.71}$$

where

$$\omega_c = \left(\frac{1}{\mu\varepsilon}\right)^{1/2}\left[\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2\right]^{1/2} = 2\pi f_c \tag{18.72}$$

This critical value of the frequency is called the cut-off frequency [see Eq. (18.51)].

The corresponding cut-off wavelength

$$\lambda_c = \frac{\mathrm{u}}{f_c}\left(= \frac{\mathrm{c}}{f_c}\right) = \frac{2\pi}{\left[\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2\right]^{1/2}} \tag{18.73}$$

The velocity of wave propagation

$$\mathrm{v} = \frac{\omega}{\beta} = \frac{\omega}{\left[\omega^2\mu\varepsilon - \left\{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2\right\}\right]^{1/2}} \tag{18.74}$$

which indicates that the velocity of propagation of the wave in the guide is greater than the phase velocity in free space. As the frequency is increased above the cut-off frequency, the phase velocity decreases from an infinitely large value and approaches the velocity of light in free space c, as the frequency is kept on increasing indefinitely.

The cut-off frequencies of different modes are different. The mode TE_{10} for which $m = 1$, $n = 0$, has the lowest cut-off frequency, i.e.

$$(f_c)_{\text{TE10}} = \frac{1}{2a\sqrt{\mu\varepsilon}} = \frac{\text{c}}{2a}, \text{ for } a >> b \tag{18.75}$$

which matches with the value obtained in Eq. (18.51).

The next lowest cut-off frequency is for TE_{01} mode, for which

$$(f_c)_{\text{TE01}} = \frac{\text{c}}{2b} \tag{18.76}$$

$\therefore$ In the range $(\text{c}/2a) < f < (\text{c}/2b)$, only TE_{10} mode can propagate through the guide.

The field equations for the TE_{mn} mode are:

$$E_x = j\left(\frac{\omega\mu}{k^2}\right)H_0 \frac{n\pi}{b}\cos\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right) \tag{18.77a}$$

$$E_y = -j\left(\frac{\omega\mu}{k^2}\right)H_0 \frac{m\pi}{a}\sin\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi y}{b}\right) \tag{18.77b}$$

$$H_x = j\left(\frac{\beta}{k^2}\right)H_0 \frac{m\pi}{a}\sin\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi y}{b}\right) \tag{18.77c}$$

$$H_y = j\left(\frac{\beta}{k^2}\right)H_0 \frac{n\pi}{b}\cos\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right) \tag{18.77d}$$

$$H_z = H_0 \cos\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi y}{b}\right) \tag{18.77e}$$

In the above expressions, γ has been replaced by $j\beta$, which is valid for the frequencies above the cut-off frequency f_c.

Implicit in the above expressions are the terms $[\exp(-j\beta z)\exp(j\omega t)]$, and the real part of the total expression has to be considered. Also, by convention, the x-coordinate is assumed to coincide with the larger transverse dimension, and hence the TE_{10} mode has the lowest cut-off frequency (lower than the TE_{01} mode), and this is called the 'dominant mode'.

18.3.3.2 *Energy transmission in the TE_{10} mode of rectangular waveguides*

The field components in the TE_{10} mode are obtained by substituting $m = 1$, $n = 0$, in Eqs. (18.77), which give:

$$E_x = 0 \tag{18.78a}$$

$$E_y = -j\left(\frac{\omega\mu a}{\pi}\right)H_0 \sin\left(\frac{\pi x}{a}\right)\exp(-j\beta z) \tag{18.78b}$$

$$H_x = j\left(\frac{\beta a}{\pi}\right)H_0 \sin\left(\frac{\pi x}{a}\right)\exp(-j\beta z) \tag{18.78c}$$

$$H_y = 0 \tag{18.78d}$$

$$H_z = H_0 \cos\left(\frac{\pi x}{a}\right) \exp\left(-j\beta z\right) \tag{18.78e}$$

All these expressions have exp ($j\omega t$) implicit in them.

$$\beta = \sqrt{\omega^2 \mu\varepsilon - \left(\frac{\pi}{a}\right)^2} = \omega\sqrt{\mu\varepsilon}\sqrt{1 - \frac{\left(\frac{\pi}{a}\right)^2}{\omega^2\mu\varepsilon}}$$

$$= \omega\sqrt{\mu\varepsilon}\sqrt{1 - \left(\frac{\mathrm{c}}{2af}\right)^2} = \omega\sqrt{\mu\varepsilon}\sqrt{1 - \left(\frac{f_c}{f}\right)^2} \tag{18.79}$$

where $\omega = 2\pi f$; $\mathrm{c} = \sqrt{\mu\varepsilon}$ and $f_c = \dfrac{\mathrm{c}}{2a} = \dfrac{1}{2a\sqrt{\mu\varepsilon}}$ both in free space.

The wavelength of the wave propagating along the z-direction of the guide is (= λ_z):

$$\lambda_z = \frac{2\pi}{\beta} = \frac{2\pi\mathrm{c}}{2\pi f\sqrt{1 - \left(\frac{f_\mathrm{c}}{f}\right)^2}} = \frac{\mathrm{c}}{f\sqrt{1 - \left(\frac{f_\mathrm{c}}{f}\right)^2}}$$

$$= \frac{\lambda}{\sqrt{1 - \frac{f_c^2}{f^2}}} \tag{18.80}$$

λ, being the wavelength of the plane wave of the same frequency.

Hence the velocity of propagation of the wave along the waveguide (in the z-direction) is:

$$\mathrm{v_p} = f\lambda_z = \frac{\mathrm{c}}{\sqrt{1 - \frac{f_c^2}{f^2}}} \tag{18.81}$$

This velocity of propagation as described before is greater than the velocity of light in vacuum. It is only a 'geometrical velocity', and the energy propagation is at a lower velocity which is the group velocity. The average power transmission through the guide is obtained by integrating the complex Poynting vector, over a cross-section of the guide. The complex Poynting vector is in the direction of propagation, i.e. in the z-direction.

$$\mathbf{S}_\mathrm{av} = \left(\frac{1}{2}\right)\mathbf{E}_\mathrm{transverse} \times \mathbf{H}^*_\mathrm{transverse} = -\left(\frac{1}{2}\right)E_y H_x^* \mathbf{i}_z$$

$$\therefore \quad |\mathbf{S}_\mathrm{av}| = \int_0^a \int_0^b \mathrm{Re}\left[\left(\frac{1}{2}\right)E_y H_x^*\right] dx\,dy$$

$$= \frac{\omega\mu\beta a^2 H_0^2}{2\pi^2} \int_0^a \int_0^b \sin^2\left(\frac{\pi x}{a}\right) dx\,dy$$

$$= \frac{\omega\mu\beta a^3 b H_0^2}{4\pi^2} \tag{18.82}$$

Typical dimensions of a guide to transmit a TE_{10} mode with a free space wavelength of 10 cm would be a = 7.5 cm and b = 2.5 cm.

18.3.3.3 Transverse magnetic wave in rectangular waveguides

If now in the rectangular waveguide $H_z = 0$, then the magnetic field is in the transverse plane, and such waves are called the transverse magnetic or TM waves.

When $H_z = 0$, Eqs. (18.53) become:

$$\gamma H_y = j\omega\varepsilon E_x \tag{18.83a}$$

$$\gamma H_x = -j\omega\varepsilon E_y \tag{18.83b}$$

$$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = j\omega\varepsilon E_z \tag{18.83c}$$

$$\frac{\partial E_z}{\partial y} + \gamma E_y = -j\omega\mu H_x \tag{18.83d}$$

$$\frac{\partial E_z}{\partial x} + \gamma E_x = j\omega\mu H_y \tag{18.83e}$$

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = 0 \tag{18.83f}$$

This time we express the field components in terms of E_z.

Combining Eqs. (18.83d) and (18.83b),

$$\frac{\partial E_z}{\partial y} + \gamma\left(\frac{-\gamma}{j\omega\varepsilon}\right)H_x = -j\omega\mu H_x$$

$$\therefore \quad H_x = \left(\frac{j\omega\varepsilon}{k^2}\right)\left(\frac{\partial E_z}{\partial y}\right) \tag{18.84}$$

Similarly from Eqs. (18.83e) and (18.83a),

$$H_y = -\left(\frac{j\omega\varepsilon}{k^2}\right)\left(\frac{\partial E_z}{\partial x}\right) \tag{18.85}$$

Combining these two equations with (18.83a) and (18.83b),

$$E_x = -\left(\frac{\gamma}{k^2}\right)\left(\frac{\partial E_z}{\partial x}\right) \tag{18.86}$$

$$E_y = -\left(\frac{\gamma}{k^2}\right)\left(\frac{\partial E_z}{\partial y}\right) \tag{18.87}$$

and the equation for E_z is

$$\frac{\partial^2 E_z}{\partial x^2} + \frac{\partial^2 E_z}{\partial y^2} + k^2 E_z = 0 \tag{18.88}$$

The method of solving this equation is same as that of Eq. (18.63). Using the same method of separation of variables, the solution can be written as

$$E_z = \sum_{k_x}\sum_{k_y}[C_x \sin(k_x x) + D_x \cos(k_x x)][C_y \sin(k_y y) + D_y \cos(k_y y)] \tag{18.89}$$

Applying the boundary conditions:

(1) On $y = 0$ and $y = b$, $E_x = 0$, $E_z = 0$.

From Eq. (18.89):

$$(E_z)_{y=0} = \Sigma\Sigma\,[C_x \sin(k_x x) + D_x \cos(k_x x)]D_y = 0 \text{ for all } x.$$

$\therefore$ $D_y = 0$; and from $y = b$ plane, $\sin(k_y b) = \sin n\pi$

$\therefore$ $k_y = \left(\frac{n\pi}{b}\right)$, where $n = 1, 2, 3, \ldots$.

(2) On $x = 0$ and $x = a$, $E_z = 0$.

This condition gives: $D_x = 0$; and $\sin(k_x a) = \sin m\pi$, $m = 1, 2, 3, \ldots$. Note that in this case $m = 0$ and $n = 0$ are not admissible values.

$\therefore$ The final expression for E_z is

$$E_z = E_0 \sin\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right) \tag{18.90}$$

where m and $n = 1, 2, 3, \ldots$; $E_0 = C_x C_y$ [from Eq. (18.89)]

Note that as in the previous case,

$$k^2 = k_x^2 + k_y^2 = \left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 = \gamma^2 + \omega^2\mu\varepsilon$$

or

$$\gamma = \left[\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 - \omega^2\mu\varepsilon\right]^{1/2}$$

$$= \alpha + j\beta \tag{18.91}$$

where α and β have the same meanings ascribed to them in Section 18.3.3.1. Since, if γ was real, β the phase-shift constant would be zero and there would be no wave in the guide. There would be, as before, a critical value of ω $(= \omega_c)$ for which γ is zero, and above which γ would be imaginary, i.e. $= j\beta$ and there would be wave propagation in the guide. This value of ω would give the cut-off frequency, whose expression is same as for the TE mode, i.e.

$$\omega_c = \left(\frac{1}{\sqrt{\mu\varepsilon}}\right)\sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2} = 2\pi f_c \tag{18.92}$$

m and n being $1, 2, 3, \ldots$.

$$\therefore \quad f_c = \left(\frac{1}{2\pi\sqrt{\mu\varepsilon}}\right)\sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2} \tag{18.93}$$

and the corresponding cut-off wavelength is

$$\lambda_c = \frac{2\pi}{\sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2}} \tag{18.94}$$

The wave number β for propagation in the TM mode is

$$\beta = \left[\omega^2 \mu\varepsilon - \left\{ \left(\frac{m\pi}{a} \right)^2 + \left(\frac{n\pi}{b} \right)^2 \right\} \right]^{1/2} \tag{18.95}$$

The velocity of wave propagation will be:

$$\mathrm{v} = \frac{\omega}{\beta} = \frac{\omega}{\left[\omega^2 \mu\varepsilon - \left\{ \left(\frac{m\pi}{a} \right)^2 + \left(\frac{n\pi}{b} \right)^2 \right\} \right]^{1/2}} \tag{18.96}$$

This is again greater than the phase velocity in free space; and its behaviour pattern is similar to that in the TE mode. The wavelength in the guide is

$$\lambda = \frac{\mathrm{v}}{f} = \frac{2\pi}{\left[\omega^2 \mu\varepsilon - \left\{ \left(\frac{m\pi}{a} \right)^2 + \left(\frac{n\pi}{b} \right)^2 \right\} \right]^{1/2}} \tag{18.97}$$

which is also greater than the free space wavelength. The lowest value for the TM mode is $m = n = 1$, i.e. TM_{11}. The expressions for the field vectors of the TM_{mn} mode are as given below:

$$E_x = -j \left(\frac{\beta}{k^2} \right) E_0 \frac{m\pi}{a} \cos \left(\frac{m\pi x}{a} \right) \sin \left(\frac{n\pi y}{b} \right) \tag{18.98a}$$

$$E_y = -j \left(\frac{\beta}{k^2} \right) E_0 \frac{n\pi}{b} \sin \left(\frac{m\pi x}{a} \right) \cos \left(\frac{n\pi y}{b} \right) \tag{18.98b}$$

$$E_z = E_0 \sin \left(\frac{m\pi x}{a} \right) \sin \left(\frac{n\pi y}{b} \right) \tag{18.98c}$$

$$H_x = j \left(\frac{\omega\varepsilon}{k^2} \right) E_0 \frac{n\pi}{b} \sin \left(\frac{m\pi x}{a} \right) \cos \left(\frac{n\pi y}{b} \right) \tag{18.98d}$$

$$H_y = -j \left(\frac{\omega\varepsilon}{k^2} \right) E_0 \frac{m\pi}{a} \cos \left(\frac{m\pi x}{a} \right) \sin \left(\frac{n\pi y}{b} \right) \tag{18.98e}$$

In these expressions, γ again has been replaced by $j\beta$ which is valid for the frequencies above the cut-off frequency. And also the terms [exp $(-j\beta z)$ exp $(j\omega t)$] are implicit in these expressions.

18.3.4 Phase Velocity and Group Velocity

During our discussion of wave propagation of guided waves between parallel planes and in waveguides, we have mentioned two types of velocities. The first one is the 'phase velocity' $(= v_{ph})$ which is the velocity of propagation of equiphase surfaces along the guide. (It should be noted that the phase velocity is identical with the velocity of propagation of the wave-fronts in the

rigorous sense only when we are dealing with non-dispersive waves). The geometrical significance of the phase velocity is shown on the ω–k plot of Figure 18.13(a), i.e.

$$v_{ph} = \frac{\omega}{k}$$

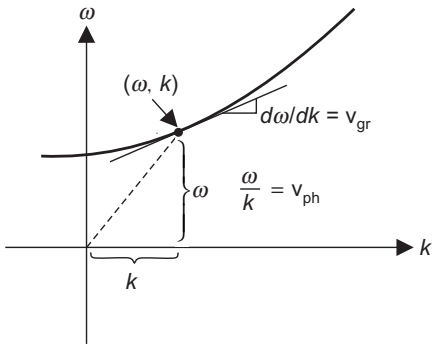


Figure 18.13(a) The slope of the line joining the point (w, k) and the origin is the phase velocity. The tangent of the w-k curve at this same point is the group velocity.

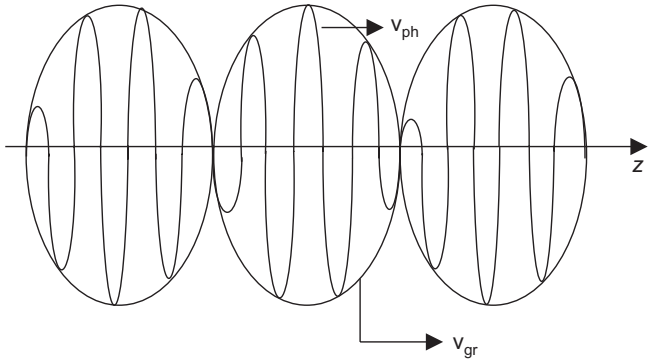


Figure 18.13(b) Sum of two sinusoidal waves of different frequencies.

For evanescent waves, the phase velocity approaches infinity as $k \to 0$, and it is certainly unreasonable to assume that a pulse will propagate with this velocity.

The second type of velocity we have come across is the group velocity ($= v_{gr}$) which in these particular cases, could be taken as the velocity of energy propagation in the direction of the axis of the guide. In general, the group velocity of the waves with frequencies in the neighbourhood of ω is defined as, $v_{gr} = (d\omega/dk)$, or as the slope shown in Figure 18.13(a).

Its physical significance is then the velocity of propagation of a group of waves with essentially the same wave-number and frequency.

For waveguide propagation, the phase velocity v_{ph} is always greater than $v_0 \left[= \sqrt{(\mu\varepsilon)^{-1/2}}\right]$, whereas the group velocity is always less than v_0.

It should be clearly understood that by itself, a purely sinusoidal wave cannot transmit any information other than its own presence. If a signal has to be transmitted, then the wave must be suitably modified. Whatever the method, the result is a 'package of frequencies' which is close to the fundamental frequency (i.e. the frequency of the wave without a signal). The velocity of the package carrying the information is the main interest.

We shall try to explain this by considering the simplest possible package made up of only two waves of angular frequencies ω_1 and ω_2 which are close to the fundamental frequency ω. Since the frequencies are different, their phase constants, (say) β_1 and β_2 respectively, would also be different. So we add the two waves, in the forms of:

$$F_1(t, z) = \cos(\omega_1 t - \beta_1 z) \qquad \text{and} \qquad F_2(t, z) = \cos(\omega_2 t - \beta_2 z)$$

By standard trigonometrical relation, which gives the sum as

$$F_1(t, z) + F_2(t, z) = 2\cos\left[\frac{(\omega_1 + \omega_2)t}{2} - \frac{(\beta_1 + \beta_2)z}{2}\right]\cos\left[\frac{(\omega_2 - \omega_1)t}{2} - \frac{(\beta_2 - \beta_1)z}{2}\right]$$

Let ω_1 and ω_2 be very near the fundamental frequency $\omega\,(\omega_2 > \omega > \omega_1)$, so that:

$$(\omega_2 + \omega_1)/2 = \omega, \quad (\beta_2 + \beta_1)/2 = \beta, \quad (\omega_2 - \omega_1) = \Delta\omega, \quad \text{and } (\beta_2 - \beta_1) = \Delta\beta.$$

The above sum of these two waves then can be expressed as

$$F_1(t, z) + F_2(t, z) = 2\cos\left(\frac{\Delta\omega t - \Delta\beta z}{2}\right)\cos(\omega t - \beta z) \tag{18.99}$$

Since $\Delta\omega << \omega$, this expression represents a signal in the form of a wave $\cos(\omega t - \beta z)$ with a varying amplitude $2\cos[(\Delta\omega t - \Delta\beta z)/2]$ as shown in Figure 18.13(b).

The sinusoidal wave inside the envelope propagates with the velocity $v_{ph} = \omega/\beta$. The wave packages enclosed by the envelope are the carriers of energy, since no energy can be transmitted through the zero values of the wave. The velocity of the envelope is $(\Delta\omega/\Delta\beta)$, which in the limiting case becomes:

$$v_{gr} = \frac{d\omega}{d\beta} = \frac{1}{\dfrac{d\beta}{d\omega}} \tag{18.100}$$

which is the group velocity in a dispersive medium. This assumes that the frequency spectrum of the signal is quite narrow. The group velocity is therefore interpreted as the velocity at which the group of waves, defined by the envelope, propagates in the z-direction. Note that, because the phase and the group velocities are different, the waves will appear to move with respect to the envelope.

If we consider the rectangular waveguide in TE_{10} mode, the phase constant was obtained as:

$$\beta = \omega\left[\sqrt{\mu\varepsilon}\sqrt{1 - \left(\frac{f_c}{f}\right)^2}\right] \qquad \text{[Eq. (18.79)]}$$

Using Eq. (18.100), the signal velocity in the guide is

$$v_{gr} = \frac{1}{\sqrt{\mu\varepsilon}}\sqrt{1 - \left(\frac{f_c}{f}\right)^2} = c\left[\sqrt{1 - \left(\frac{f_c}{f}\right)^2}\right] \tag{18.101}$$

and, in this case only:

$$v_{gr}v_{ph} = c^2 \tag{18.102}$$

18.3.5 Rectangular Waveguide: Attenuation due to Lossy Dielectric, and Imperfectly Conducting Walls

We shall briefly recapitulate the wave behaviour in imperfect conductors and lossy dielectrics.

Maxwell's equations become:

$$\nabla \times \mathbf{E} = -\omega\mu\mathbf{H} \tag{18.103a}$$

$$\nabla \times \mathbf{H} = (\sigma + j\omega\varepsilon)\mathbf{E} = j\omega\varepsilon\left(1 + \frac{\sigma}{j\omega\varepsilon}\right)\mathbf{E} \tag{18.103b}$$

$\therefore$ We have to replace ε by $[\varepsilon\{1 - j\sigma/(\omega\varepsilon)\}]$ in all the formulae for propagation, impedance, etc.; e.g.

$$\gamma\,(= j\beta) = j\omega\sqrt{(\mu\varepsilon)\left(1 + \frac{j\sigma}{\omega\varepsilon}\right)}, \text{ which is complex} \tag{18.104}$$

$$= \alpha + j\beta$$

$$Z_0 = \left[\frac{\mu}{\varepsilon\left(1 - \frac{j\sigma}{\omega\varepsilon}\right)}\right]^{1/2} \leftarrow \text{characteristic impedance of the medium} \tag{18.105}$$

$$\frac{\sin\theta_r}{\sin\theta_i} = \left[\frac{\mu_1\varepsilon_1}{\mu_2\varepsilon_2\left(1 - \frac{j\sigma}{\omega\varepsilon_2}\right)}\right]^{1/2} \tag{18.106}$$

The critical quantity in this expression is $\sigma/\omega\varepsilon_2$.

For a good conductor, $\sigma/\omega\varepsilon_2 >> 1$,

$$\frac{\sin\theta_r}{\sin\theta_i} = \left[\frac{\mu_1\varepsilon_1}{\frac{\mu_2\sigma}{j\omega}}\right]^{1/2} = \left(\frac{j\omega\varepsilon_1}{\sigma}\right)^{1/2} \tag{18.107}$$

which is very small.

$\therefore$ $\sin\theta_r$ is very small for any value of θ_i, i.e. the propagation takes place almost normally into the conducting medium. Also, since the above expression is complex, the wave gets exponentially damped as it propagates.

18.3.5.1 Attenuation due to lossy dielectric

We consider the TE_{10} mode in the rectangular waveguide. For the loss-less dielectric, the propagation constant γ is

$$\gamma = j\beta = jk\left[\sqrt{1 - \left(\frac{f_c}{f}\right)^2}\right] = j\omega\left(\sqrt{\mu\varepsilon}\right)\left[\sqrt{1 - \left(\frac{f_c}{f}\right)^2}\right] \tag{18.108}$$

where $k^2 = \omega^2(\mu\varepsilon)$.

For the lossy dielectric, we have replaced

$$\varepsilon \text{ by } \varepsilon\left(1 - \frac{j\sigma}{\omega\varepsilon}\right) \tag{18.109}$$

$\therefore$ k gets modified as shown below, i.e.

$$k = \omega\left(\sqrt{\mu\varepsilon}\right)\sqrt{1 + \frac{\sigma}{j\omega\varepsilon}} = \omega\left(\sqrt{\mu\varepsilon}\right)\left(1 + \frac{\sigma}{j2\omega\varepsilon}\right) \tag{18.110}$$

$\therefore$ The propagation constant is now complex, i.e.

$$\gamma = \alpha + j\beta = j\omega\left(\sqrt{\mu\varepsilon}\right)\left(1 + \frac{\sigma}{j2\omega\varepsilon}\right)\sqrt{1 - \left(\frac{f_c}{f}\right)^2} \tag{18.111}$$

$\therefore$ The attenuation constant

$$\alpha = \omega\left(\sqrt{\mu\varepsilon}\right)\frac{\sigma}{2\omega\varepsilon}\sqrt{1 - \left(\frac{f_c}{f}\right)^2}$$

$$= \left(\sqrt{\frac{\mu}{\varepsilon}}\right)\left(\frac{\sigma}{2}\right)\sqrt{1-\left(\frac{f_c}{f}\right)^2} \qquad \text{Nepers/metre} \tag{18.112}$$

and the phase-shift constant

$$\beta = \omega\left(\sqrt{\mu\varepsilon}\right)\sqrt{1-\left(\frac{f_c}{f}\right)^2} \tag{18.113}$$

18.3.5.2 Attenuation due to imperfectly conducting walls

Let us initially consider transmission lines with sending waves (V_S, I_S):

$$V = V_S \exp(-\alpha z), \qquad I = I_S \exp(-\alpha z) \tag{18.114}$$

$$\therefore \quad \text{Transmitted power} = VI\cos\phi = V_S I_S \cos\phi \exp(-2\alpha z) = W \tag{18.115}$$

$$\therefore \quad \frac{dW}{dz} = -2\alpha W \tag{18.116}$$

$$\therefore \quad \alpha = \frac{\text{Power loss per unit length}}{2 \times \text{Power sent from the sending point}} \tag{18.117}$$

$$\text{Power dissipated} = \left(\frac{1}{2}\right)\text{Re}(\mathbf{E}_t \times \mathbf{H}_t^*) = \left(\frac{1}{2}\right)\text{Re}(\mathbf{H}_t Z_{in} \times \mathbf{H}_t^*)$$

$$= \left(\frac{1}{2}\right)\text{Re}\left(|\mathbf{H}_t|^2 Z_{in}\right)$$

$$= |\mathbf{H}_t|^2 \left(\frac{1}{2}\right)\text{Re}\sqrt{\frac{j\omega\mu}{\sigma}}$$

(see below for this step)

$$= |\mathbf{H}_t|^2 \left(\frac{1}{2}\right)\sqrt{\frac{\omega\mu}{\sigma}} = \left(\frac{1}{2}\right)|\mathbf{J}|^2 R_S \tag{18.118}$$

Note: To calculate the input impedance due to (say) the bottom wall (of the waveguide),

$$Z_{in} = \frac{E_t}{H_t} = \left[\frac{\mu}{\varepsilon\left(1-\frac{j\sigma}{\omega\varepsilon}\right)}\right]^{1/2} = \sqrt{\frac{j\omega\mu}{\sigma}} \tag{18.119}$$

The tangential E_t exists on the wall because now there is an ohmic loss in the wall as it is now an imperfect conductor.

Also, when there is a surface current, $H_{t1} - H_{t2} = J_S$.

For a perfect conductor, $H_{t2} = 0$, and hence $H_{t1} = J_S$ (18.120)

Referring to Figure 18.12, we consider first the top and the bottom walls, which are planes $y = 0$ and $y = b$. We write the field vectors [from Eqs. (18.78)] as:

$$E_y = A_0 \sin\left(\frac{\pi x}{a}\right) \exp(-\gamma z) \tag{18.121a}$$

$$H_x = -\left(\frac{\gamma}{\omega\mu}\right) A_0 \sin\left(\frac{\pi x}{a}\right) \exp(-\gamma z) \tag{18.121b}$$

$$H_z = -\left(\frac{\pi}{j\omega\mu a}\right) A_0 \cos\left(\frac{\pi x}{a}\right) \exp(-\gamma z) \tag{18.121c}$$

$\therefore$ On the top and the bottom walls, $y = 0$ and $y = b$,

$$J_z = -\left(\frac{\gamma}{\omega\mu}\right) A_0 \sin\left(\frac{\pi x}{a}\right) \exp(-\gamma z) \tag{18.122a}$$

$$J_x = -\left(\frac{\pi}{j\omega\mu a}\right) A_0 \cos\left(\frac{\pi x}{a}\right) \exp(-\gamma z) \tag{18.122b}$$

$\therefore$ The total power loss in the top + bottom plates per unit length in the z-direction is

$$= \int_0^a |J|^2 R_S dx \cdot 2 = \int_0^a \left(|J_z|^2 + |J_x|^2\right) R_S dx \cdot 2$$

$$= \left[\left(\frac{\beta^2}{\omega^2\mu^2}\right) A_0^2 \left(\frac{a}{2}\right) + \left(\frac{\pi^2}{\omega^2\mu^2 a^2}\right) A_0^2 \left(\frac{a}{2}\right)\right] R_S \cdot 2 \tag{18.123}$$

Next, we consider the side walls $x = 0$ and $x = a$; on these side walls, since $H_y = 0$, only J_y will be produced by H_z, i.e.

$$J_y = -\left(\frac{\pi}{j\omega\mu a}\right) A_0 \cos\left(\frac{\pi x}{a}\right) \exp(-\gamma z) \tag{18.124}$$

$\therefore$ The power loss in the two side plates per unit length

$$= \left(\frac{\pi^2}{\omega^2\mu^2 a^2}\right) A_0^2 \, a R_S \tag{18.125}$$

The transmitted power through the guide

$$= \int_0^a \int_0^b \left(\frac{1}{2}\right) \mathrm{Re}\left(\mathbf{E}_y \times \mathbf{H}_x^*\right) dx dy$$

$$= \int_0^a \left(\frac{j\gamma}{\omega\mu}\right) A_0^2 \sin^2\left(\frac{\pi x}{a}\right) dx \tag{18.126}$$

$$= \left(\frac{j\gamma}{\omega\mu}\right) A_0^2 \, b \left(\frac{a}{2}\right)$$

Note: $\exp(-2\gamma z)$ has been assumed in all these expressions, and R_S the resistance of the walls per unit length.

18.3.5.3 *Parallel plane waveguide (as a limiting case of the rectangular waveguide)*

Initially, the parallel plane waveguide was derived from the reflection of uniform plane waves hitting a perfectly conducting plane surface obliquely and studying the interference pattern obtained therefrom. The rectangular waveguide was first derived from the parallel plane waveguide by adding two parallel conducting surfaces orthogonally. The rectangular waveguide was also analysed by using general mathematical approach, from fundamental considerations. Now, we shall treat the parallel plane waveguide as a limiting case of the rectangular waveguide.

A rectangular waveguide with its edges parallel to the coordinate axes is shown in Figure 18.13(c) (as was shown in Figure 18.12 as well). For the rectangular waveguide, with its dimensions as shown in Figure 18.13(c), we have

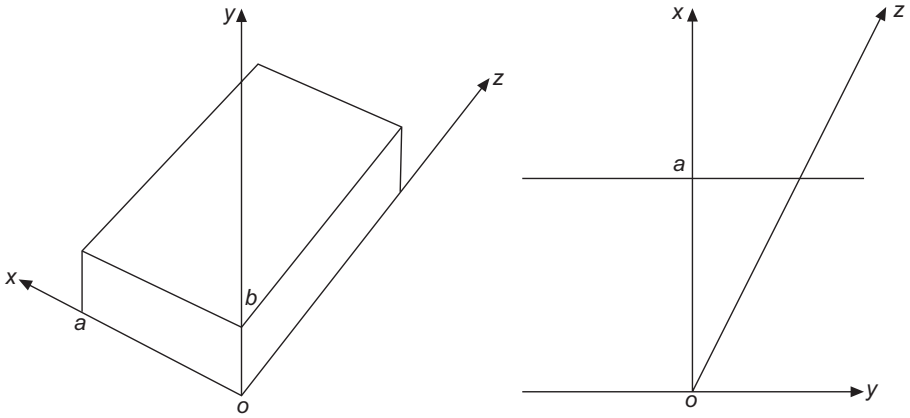


Figure 18.13(c) Rectangular waveguide ($a \times b$) and its limit as $b \to \infty$ giving the plane parallel waveguide.

$$\gamma = \left[\left(\frac{m\pi}{a} \right)^2 + \left(\frac{n\pi}{b} \right)^2 - \omega^2 \mu\varepsilon \right]^{1/2} \tag{18.70}$$

When $\omega > \omega_c$,

$$\gamma = j\beta = j \left[\omega^2 \mu\varepsilon - \left\{ \left(\frac{m\pi}{a} \right)^2 + \left(\frac{n\pi}{b} \right)^2 \right\} \right]^{1/2} \tag{18.71}$$

where

$$\omega_c = \left(\frac{1}{\mu\varepsilon} \right)^{1/2} \left[\left(\frac{m\pi}{a} \right)^2 + \left(\frac{n\pi}{b} \right)^2 \right]^{1/2} = 2\pi f_c$$

To get the limit $b \to \infty$, then

$$j\beta = j \left[\omega^2 \mu\varepsilon - \left(\frac{m\pi}{a} \right)^2 \right]^{1/2} \quad \text{and} \quad k^2 = \left(\frac{m\pi}{a} \right)^2$$

and the field equations for the TE mode become [(from Eqs. (18.77)]

$$\left.\begin{aligned} E_x &= 0 \\ E_y &= -\frac{j\omega\mu}{m\pi/a} H_0 \sin\left(\frac{m\pi x}{a}\right) \exp[j(\omega t - \beta z)] \\ H_x &= \frac{j\beta}{m\pi/a} H_0 \sin\left(\frac{m\pi x}{a}\right) \exp[j(\omega t - \beta z)] \\ H_y &= 0 \\ H_z &= H_0 \cos\left(\frac{m\pi x}{a}\right) \exp[j(\omega t - \beta z)] \end{aligned}\right\} \qquad (18.77')$$

These are same as Eqs. (17.204a) of Section 18.3.1.1.

Note: In the above equation, the direction of propagation is z, whereas in Section 18.3.1.1 the direction of wave propagation is x.

Next we consider the TM mode for the rectangular waveguide. As before, we take the limit as $b \to \infty$, but in this case all the five expressions [i.e. Eq. (18.98)] become identically equal to zero. So the question that arises is 'can the fields of the TM mode be derived thus or that the TM mode does not exist. The answer to this apparent confusion lies in the fact that H_0 is an arbitrary mathematical constant and it can be infinite. The expressions appeared to become zeroes as $b \to \infty$ was because of the implicit assumption that E_0 in finite. So we define a new constant E_0' such that

$$E_0' = E_0 \sin\frac{n\pi y}{b}$$

which is finite when $b \to \infty$.

Then the field expressions of Eqs. (18.98) become:

$$\left.\begin{aligned} E_z &= E_0' \sin\left(\frac{m\pi a}{a}\right) \exp[j(\omega t - \beta z)] \\ E_x &= -\frac{j\beta}{m\pi/a} E_0' \cos\left(\frac{m\pi x}{a}\right) \exp[j(\omega t - \beta z)] \\ E_y &= 0 \\ H_x &= 0 \\ H_y &= -\frac{j\omega\mu}{m\pi/a} E_0' \cos\left(\frac{m\pi x}{a}\right) \exp[j(\omega t - \beta z)] \end{aligned}\right\} \qquad (19.98')$$

As before these are same as Eqs. (17.191′) of Section 18.3.1.1. Again it should be noted that the direction of propagation has changed from x to z which causes a change in one of the field components. The relationship between E_{0i+} and H_0 and E_{0i} and E_0' can be written down by comparing the two sets of equations.

18.3.6 Cylindrical Waveguide

We start with the solution of Maxwell's equations in cylindrical coordinates (Figure 18.14).

We consider the Maxwell's equation

$$\oint \mathbf{E} \cdot d\mathbf{l} = -\left(\frac{d\Phi}{dt}\right)$$

and apply it to the elemental contour ABCDA of the circular guide as shown in Figure 18.14(b).

$$E_r dr + \left[E_\theta + \left(\frac{\partial E_\theta}{\partial r}\right)dr\right](r + \delta r)\, d\theta - \left[E_r + \left(\frac{\partial E_r}{\partial \theta}\right)d\theta\right]dr - E_\theta r d\theta = -j\omega\mu H_z dr r d\theta$$

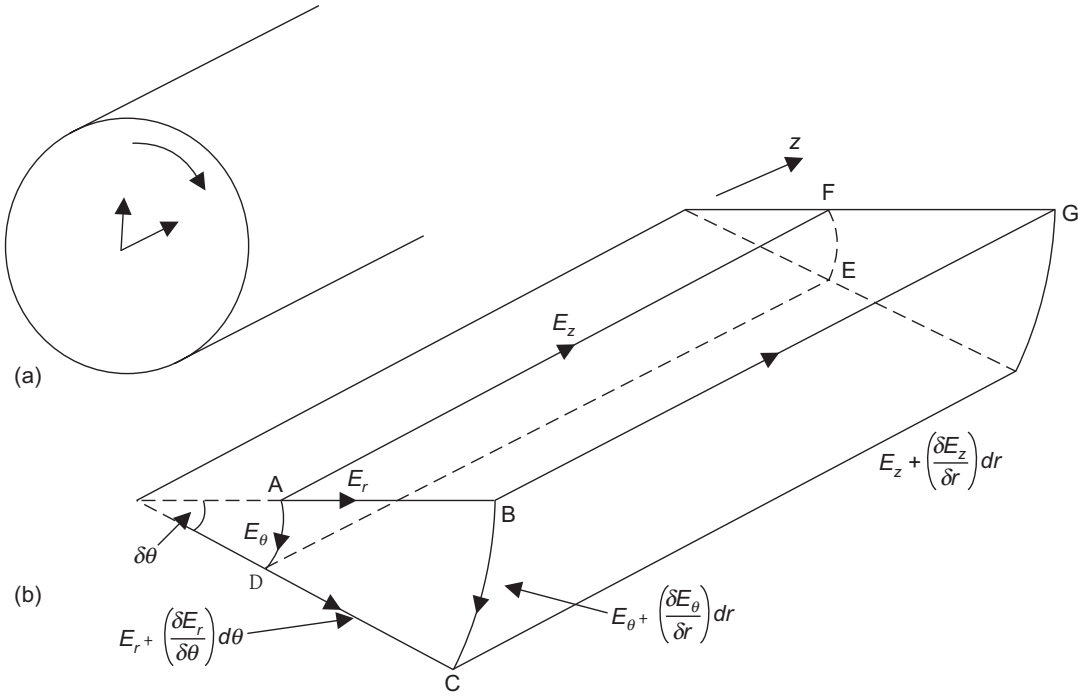


Figure 18.14 A cylindrical waveguide and its elemental section.

or

$$\frac{1}{r}\left[E_\theta + r\left(\frac{\partial E_\theta}{\partial r}\right) - \left(\frac{\partial E_r}{\partial \theta}\right)\right] = -\mu\left(\frac{\partial H_z}{\partial t}\right) \tag{18.127}$$

And similarly from the other equation $\oint \mathbf{H} \cdot d\mathbf{l} = \frac{\partial \mathbf{D}}{\partial t}$, we get

$$\frac{1}{r}\left[H_\theta + r\left(\frac{\partial H_\theta}{\partial r}\right) - \left(\frac{\partial H_r}{\partial \theta}\right)\right] = \varepsilon\left(\frac{\partial E_z}{\partial t}\right) \tag{18.128}$$

Note that $(\partial/\partial t) \equiv j\omega$, for sinusoidally time-varying excitations, and hence the operator and the imaginary quantity are interchangeable in these equations.

We again apply the line integral $\oint \mathbf{E} \cdot d\mathbf{l}$ around the contour ADEFA (in the clockwise direction) and obtain

$$E_\theta r\,dr + \left[E_z + \left(\frac{\partial E_z}{\partial \theta}\right)d\theta\right]dz - \left[E_\theta + \left(\frac{\partial E_\theta}{\partial z}\right)dz\right]rd\theta - E_z dz = -j\omega\mu H_r rd\theta dz$$

or

$$\left(\frac{1}{r}\right)\left(\frac{\partial E_z}{\partial \theta}\right) - \left(\frac{\partial E_\theta}{\partial z}\right) = -j\omega\mu H_r \tag{18.129}$$

Similarly applying $\oint \mathbf{H} \cdot d\mathbf{l}$ to the same contour ADEFA, we get

$$\left(\frac{1}{r}\right)\left(\frac{\partial H_z}{\partial \theta}\right) - \left(\frac{\partial H_\theta}{\partial z}\right) = +j\omega\varepsilon E_r \tag{18.130}$$

Next we apply $\oint \mathbf{E} \cdot d\mathbf{l}$ around the contour AFGBA (in the clockwise direction), and get

$$E_z dz + \left[E_r + \left(\frac{\partial E_r}{\partial z}\right)dz\right]dr - \left[E_z + \left(\frac{\partial E_z}{\partial r}\right)dr\right]dz - E_r dr = -j\omega\mu H_\theta dr dz$$

or

$$\frac{\partial E_r}{\partial z} - \frac{\partial E_z}{\partial r} = -j\omega\varepsilon H_\theta \tag{18.131}$$

Similarly applying $\oint \mathbf{E} \cdot d\mathbf{l}$ to the same contour AFGBA, we get

$$\frac{\partial H_r}{\partial z} - \frac{\partial H_z}{\partial r} = j\omega\varepsilon E_\theta \tag{18.132}$$

And we assume $\exp(-\gamma z)$ variation of the field in the z-direction. We then combine Eqs. (18.129) and (18.132),

$$\left(\frac{1}{r}\right)\left(\frac{\partial E_z}{\partial \theta}\right) + \gamma E_\theta = -j\omega\mu H_r$$

$$= -j\omega\mu\left[-j\left(\frac{\omega\varepsilon}{\gamma}\right)E_\theta - \left(\frac{1}{\gamma}\right)\left(\frac{\partial H_z}{\partial r}\right)\right]$$

or

$$E_\theta\left(\gamma^2 + \omega^2\mu\varepsilon\right) = -\left(\frac{\gamma}{r}\right)\left(\frac{\partial E_z}{\partial \theta}\right) + j\omega\mu\left(\frac{\partial H_z}{\partial r}\right)$$

$$\therefore \qquad E_\theta = \left(\frac{1}{k^2}\right)\left[-\left(\frac{\gamma}{r}\right)\left(\frac{\partial E_z}{\partial \theta}\right) + j\omega\mu\left(\frac{\partial H_z}{\partial r}\right)\right] \tag{18.133}$$

where $k^2 = \gamma^2 + \omega^2\mu\varepsilon$.

From Eq. (18.129),

$$H_r = -\left(\frac{1}{j\omega\mu r}\right)\left(\frac{\partial E_z}{\partial\theta}\right) - \left(\frac{\gamma}{j\omega\mu}\right)\left[-\left(\frac{\gamma}{k^2 r}\right)\left(\frac{\partial E_z}{\partial\theta}\right) + j\left(\frac{\omega\mu}{k^2}\right)\left(\frac{\partial H_z}{\partial r}\right)\right]$$

or

$$H_r = \frac{1}{k^2}\left[j\left(\frac{\omega\varepsilon}{r}\right)\left(\frac{\partial E_z}{\partial\theta}\right) - \gamma\left(\frac{\partial H_z}{\partial r}\right)\right] \tag{18.134}$$

From Eqs. (18.130) and (18.131), we get

$$H_\theta = \frac{1}{k^2}\left[-j\omega\varepsilon\left(\frac{\partial E_z}{\partial r}\right) - \left(\frac{\gamma}{r}\right)\left(\frac{\partial H_z}{\partial\theta}\right)\right] \tag{18.135}$$

$$E_r = \frac{1}{k^2}\left[-j\left(\frac{\omega\mu}{r}\right)\left(\frac{\partial H_z}{\partial\theta}\right) - \gamma\left(\frac{\partial E_z}{\partial r}\right)\right] \tag{18.136}$$

We now derive the wave equation for E_z, i.e. from Eq. (18.128),

$$\frac{1}{r}\left[H_\theta + r\left(\frac{\partial H_\theta}{\partial r}\right) - \left(\frac{\partial H_r}{\partial\theta}\right)\right] = j\omega\varepsilon E_z$$

From Eq. (18.135),

$$\frac{\partial H_\theta}{\partial r} = \frac{1}{k^2}\left[-j\omega\varepsilon\left(\frac{\partial^2 E_z}{\partial r^2}\right) + \left(\frac{\gamma}{r^2}\right)\left(\frac{\partial H_z}{\partial\theta}\right) - \left(\frac{\gamma}{r}\right)\left(\frac{\partial^2 H_z}{\partial\theta\partial r}\right)\right]$$

Also from Eq. (18.134),

$$H_r = j\left(\frac{\omega\varepsilon}{k^2 r}\right)\left(\frac{\partial^2 E_z}{\partial\theta^2}\right) - \left(\frac{\gamma}{k^2}\right)\left(\frac{\partial^2 H_z}{\partial r\partial\theta}\right)$$

Substituting from these equations in Eq. (18.128), we have

$$-j\left(\frac{\omega\varepsilon}{k^2}\right)\left(\frac{\partial E_z}{\partial r}\right) - \left(\frac{\gamma}{rk^2}\right)\left(\frac{\partial H_z}{\partial\theta}\right) - j\left(\frac{\omega\varepsilon r}{k^2}\right)\left(\frac{\partial^2 E_z}{\partial r^2}\right) + \left(\frac{\gamma}{k^2 r}\right)\left(\frac{\partial H_z}{\partial\theta}\right) - \left(\frac{\gamma}{k^2}\right)\left(\frac{\partial^2 H_z}{\partial\theta\partial r}\right)$$

$$-j\left(\frac{\omega\varepsilon}{k^2 r}\right)\left(\frac{\partial^2 E_z}{\partial\theta^2}\right) + \left(\frac{\gamma}{k^2}\right)\left(\frac{\partial^2 H_z}{\partial r\partial\theta}\right) = j\omega\varepsilon r E_z$$

Simplifying

$$\left(\frac{\partial^2 E_z}{\partial r^2}\right) + \left(\frac{1}{r}\right)\left(\frac{\partial E_z}{\partial r}\right) + \left(\frac{1}{r^2}\right)\left(\frac{\partial^2 E_z}{\partial\theta^2}\right) + \left(\gamma^2 + \omega^2\mu\varepsilon\right)E_z = 0 \tag{18.137}$$

where $k^2 = \gamma^2 + \omega^2\mu\varepsilon = \gamma^2 + p^2$, $p^2 = \omega^2\mu\varepsilon$.

This is the wave equation in E_z in the cylindrical coordinate system.

To obtain the solution of this wave equation, using the method of separation of variables, let the solution be of the form

$$E_z = RF \exp(j\omega t - \gamma z) \tag{18.138}$$

where $R \equiv R(r)$, $F = F(\theta)$.

Substituting in Eq. (18.137),

$$R''F + \left(\frac{1}{r}\right) R'F + \left(\frac{1}{r^2}\right) RF'' + (\gamma^2 + p^2)\, RF = 0$$

where the prime indicates differentiation with respect to its corresponding independent variable. Multiplying this equation by $[r^2/(RF)]$, and rearranging the terms,

$$\frac{r^2 R''}{R} + \frac{rR'}{R} + (\gamma^2 + p^2) = -\frac{F''}{F} \tag{18.139}$$

For this equation to be true for all values of r and θ, both sides must equal constants. Hence,

$$\frac{d^2F}{d\theta^2} = -n^2 F \tag{18.140}$$

where the constant n^2 is to determine the periodicity in the θ direction. The solution of this equation is

$$F = A \cos n\theta + B \sin n\theta \tag{18.141}$$

where $n = 0, 1, 2, 3, \ldots$.

The corresponding r equation is

$$\left(\frac{d^2R}{dr^2}\right) + \left(\frac{1}{r}\right)\left(\frac{dR}{dr}\right) + \left(\frac{k^2 - n^2}{r^2}\right) R = 0 \tag{18.142}$$

There are three solutions to this equation, which are

$$R = CJ_n(kr) + DY_n(kr) \tag{18.143}$$

where J_n and Y_n are the Bessel's functions of first and second kind (and of the order n) respectively. This is a standing wave solution.

$$R = EH_n^{(1)}(kr) + FH_n^{(2)}(kr) \tag{18.144}$$

where $H_n^{(1)}$ and $H_n^{(2)}$ are Hankel's functions of the first and second kind (and of the order n) respectively. These are Bessel's functions with complex arguments.

$$R = GI_n(jkr) + G'K_n(jkr) \tag{18.145}$$

where I_n and K_n are modified Bessel's functions of the first and second kind (and of order n) respectively. These are Bessel's functions with imaginary arguments.

The behaviour of $J_n(x)$ and $Y_n(x)$ are shown below (approximately) in Figure 18.15.

$H_n^{(1)}(x) = J_n(x) + jY_n(x) \leftarrow$ represents an inward travelling wave.

$H_n^{(2)}(x) = J_n(x) - jY_n(x) \leftarrow$ represents an outward travelling wave.

The inward and the outward directions refer to the assumptions based on that of exp $(j\omega t)$ (this is a matter of relative convention. Stratton, in his book uses as reference exp $(-j\omega t)$, and his notations are reversed).

$H_n^{(1)}(x)$ at the Brewster's angle is a surface wave travelling inwards so that there is no reflection.

So far, we have used the notation

$$k^2 = (\gamma^2 + \omega^2\mu\varepsilon) = \gamma^2 + p^2, \qquad p^2 = \omega^2\mu\varepsilon$$
$$= p^2 - \beta^2$$

If k is real, $p^2 - \beta^2$ is +ve; or $\left(\frac{2\pi}{\lambda_0} - \frac{2\pi}{\lambda_g}\right)$ is +ve; i.e. if k is real, $\lambda_0 < \lambda_g$.

$\therefore$ It is a fast wave, i.e. v > c.

For a slow wave, k^2 is −ve; $\therefore$ $k = j$ ( ).

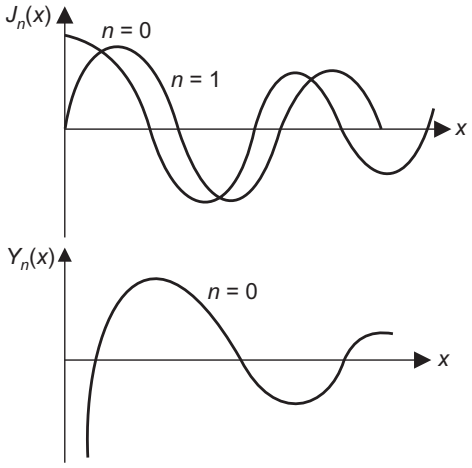


Figure 18.15 Approximate plots of $J_n(x)$ and $Y_n(x)$.

Given below in Figure 18.16 are the plots of the modified Bessel's functions which are non-orthogonal and hence are not oscillatory in nature.

$$I_n(\mu) = j^n J_n(-j\mu) \qquad \text{and} \qquad K_n(\mu) = \left(\frac{\pi}{2}\right)(-j)^{n+1} H_n^{(2)}(-j\mu)$$

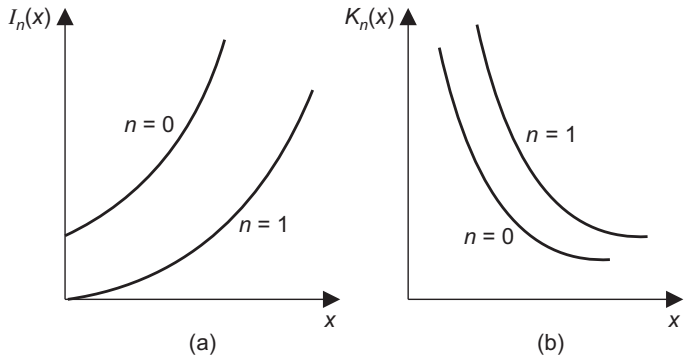


Figure 18.16 Modified Bessel's functions: (a) of the first kind and (b) of the second kind.

Next we consider some examples.

18.3.6.1 TE wave in cylindrical waveguide

We assume perfectly conducting wall; and so there is a standing wave. For the TE mode $E_z = 0$. Hence the wave equation solution is for H_z, which is

$$H_z = [CJ_n(kr) + DY_n(kr)](A \cos n\theta + B \sin n\theta) \exp(-j\beta z) \tag{18.146}$$

where $k^2 = \gamma^2 + p^2$, $\gamma = j\beta$.

Since the origin of the coordinate system is inside the guide geometry, $D = 0$ (as $Y_n \to \infty$ at $r = 0$) (see Figure 18.17).

Since the boundary conditions need the electric fields, we determine the E_θ now:

$$E_\theta = j\left(\frac{\omega\mu}{k^2}\right)\left(\frac{\partial H_z}{\partial r}\right)$$

$$= j\left(\frac{\omega\mu}{k^2}\right)kCJ_n'(kr)(A \cos n\theta + B \sin n\theta) \exp(-j\beta z) \tag{18.147}$$

where J_n' is the derivative of J_n with respect to its argument.

The boundary condition is $E_\theta = 0$ at $r = a$;

$$\therefore \quad J_n'(ka) = 0 \tag{18.148}$$

The roots of this equation are

$$(ka)'_{01} = 3.83, \qquad (ka)'_{02} = 7.02$$

$$(ka)'_{11} = 1.84, \qquad (ka)'_{12} = 5.33$$

where the first subscript refers to the value of n, and the second to the order of the root.

$$\therefore \quad \beta^2 = p^2 - k^2 = \omega^2\mu_0\varepsilon_0 - \frac{(k_{mn}a)^2}{a^2} \quad \text{(for free space)} \tag{18.149}$$

$$\therefore \quad \text{The cut-off frequency} = \omega_c^2 = \frac{(k_{mn}a)^2}{a^2\mu_0\varepsilon_0} \tag{18.150}$$

(Note that the other boundary condition that can be applied is: $H_r = 0$ at $r = a$.)

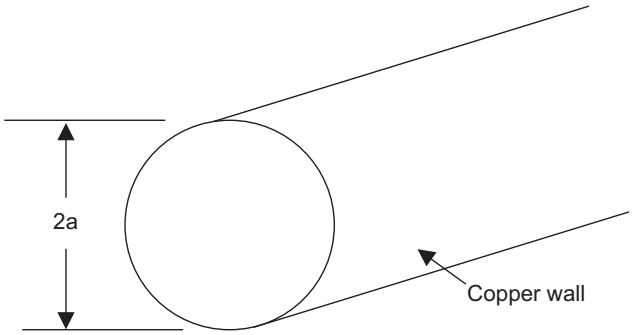


Figure 18.17 A cylindrical waveguide.

The other field components are:

$$E_r = j\left(\frac{\omega\mu n}{k^2 r}\right)A'' J_n(kr) \sin n\theta \tag{18.151a}$$

$$H_\theta = \left(\frac{n\gamma}{k^2 r}\right)A'' J_n(kr) \sin n\theta \tag{18.151b}$$

$$H_r = -\left(\frac{\gamma}{k}\right) A'' J_n'(kr) \cos n\theta \tag{18.151c}$$

where $A'' = CA$.

It should be noted that when $n = 0$ (i.e. no angular variation), the only nonzero components are E_θ, H_r, H_z. The field patterns are as shown below in Figure 18.18.

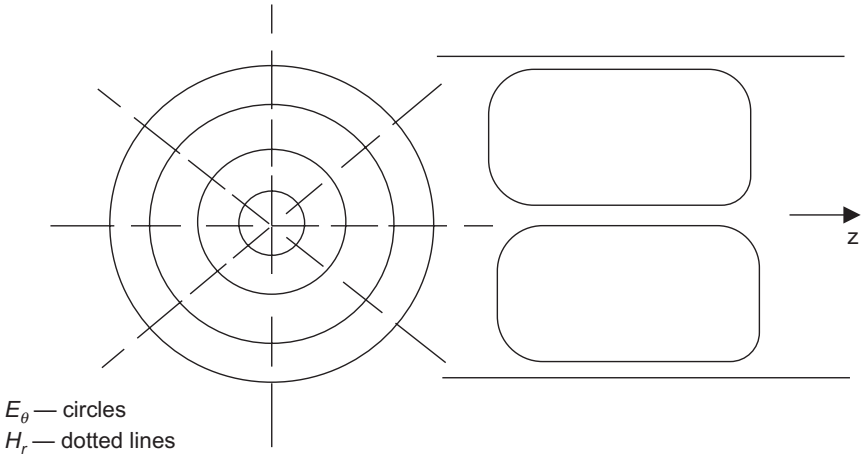


Figure 18.18 Field pattern for circular waveguide when $n = 0$.

18.3.6.2 Attenuation in TE_{10} mode in cylindrical waveguide

The field components are:

$$H_z = A'' J_0(kr) \exp(-j\beta z) \tag{18.152a}$$

$$E_\theta = A'' j\left(\frac{\omega\mu}{k}\right) J_0'(kr) \exp(-j\beta z) \tag{18.152b}$$

$$H_z = A'' j\left(\frac{\beta}{k}\right) J_0'(kr) \exp(-j\beta z) \tag{18.152c}$$

We have already shown [Eq. (18.116)] that:

$$\alpha = \frac{\text{Power lost per unit length}}{2 \times \text{Power sent}}$$

$$\text{Power sent (axially)} = \frac{1}{2}\iint \text{Re}(\mathbf{E} \times \mathbf{H}^*) dS$$

$$= \frac{1}{2}\iint \text{Re}\,(E_r H_\theta^* - E_\theta H_r^*) r\, dr\, d\theta \tag{18.153}$$

In the above expression, $E_r = 0$.

$$\therefore \quad E_\theta H_r^* = -\frac{\omega\mu\beta}{k^2}\left(A''\right)^2 \left[J_0'(kr)\right]^2$$

$$\therefore \qquad 2 \times \text{Power sent} = \int_0^{2\pi}\int_0^{a}\left(\frac{\omega\mu\beta}{k^2}\right)(A'')^2\left[J_0'(kr)\right]^2 r\, dr\, d\theta$$

$$= 2\pi\int_0^{a}\left(\frac{\omega\mu\beta}{k^2}\right)(A'')^2\left[J_0'(kr)\right]^2 r\, dr \qquad (18.154)$$

Note: $J_0' = J_1$; and

$$\int_0^{a} r\left[J_0'(kr)\right]^2 dr = \left(\frac{1}{2}\right)a^2\left[\left\{J_1(ka)\right\}^2 + \left(1 - \frac{1}{k^2a^2}\right)\left\{J_1(ka)\right\}^2\right] \qquad (18.155)$$

$$\text{Power lost/metre} = \left(\frac{1}{2}\right)J_\theta R_S \qquad (18.156)$$

where J_θ is the linear current density in amperes/metre; and

$$R_S = \sqrt{\frac{\pi\beta\mu}{\sigma}} \qquad (18.157)$$

$$\therefore \quad \text{Power lost per metre} = \left(\frac{1}{2}\right)H_z R_S = \left(\frac{1}{2}\right)R_S\,(A'')^2\,[J_0(ka)]^2\,2\pi al \qquad (18.158)$$

(l, being the length of the guide, which in this case is = 1.)

$$\therefore \qquad \alpha = \frac{dP}{2P} = \left(\frac{R_S}{aZ_0}\right)\left[\frac{1}{\sqrt{1 - \left(\frac{f_c}{f}\right)^2}}\right]\left(\frac{f_c}{f}\right)^2 \qquad (18.159)$$

If α is plotted as a function of frequency, then its behaviour is as shown in Figure 18.19, given below.

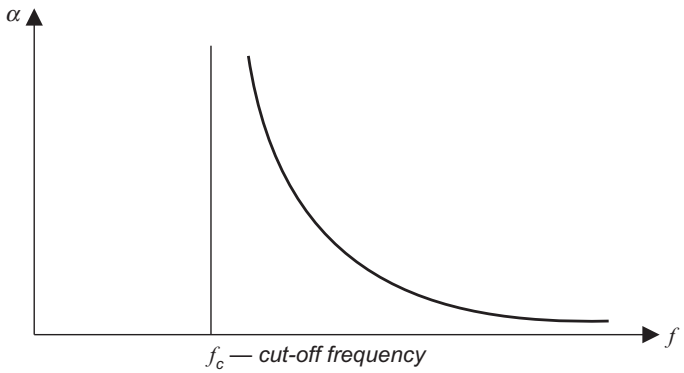


Figure 18.19 Attenuation as a function of the frequency.

We show below the two modes TE_{01} and TE_{11} with their **E** and **H** lines shown in two orthogonal sections (Figure 18.20).

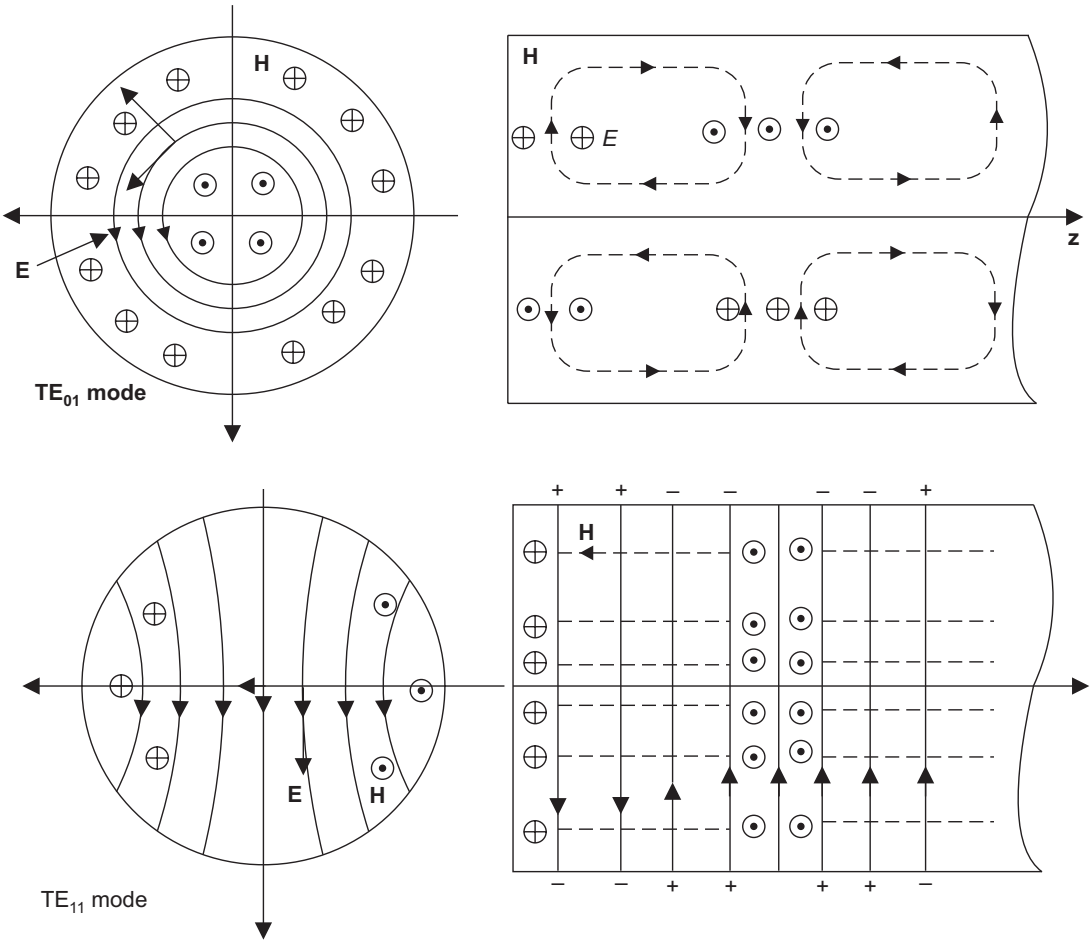


Figure 18.20 TE_{01} and TE_{11} modes in a circular waveguide.

18.3.7 Quality Factor Q of Waveguides

We have already discussed earlier, a number of times, the attenuation factor (or constant) α in the range of propagation of a wave. Another quantity closely related to α is the 'quality factor Q' which is defined as:

$$Q = \frac{\omega\,(\text{Energy stored per unit length})}{\text{Energy lost per unit length per second}} \tag{18.160}$$

$$= -\omega W\left(\frac{\partial W}{\partial t}\right)^{-1} = \frac{\omega W}{\rho}$$

where W = energy at a time after the initial energy W_0.

The above expression, which is a general one, is applicable to both waveguides and transmission lines.

For a waveguide,

$$\text{Energy transmitted per second} = v_{gr} \times \text{Energy stored per unit length} \tag{18.161}$$

where v_{gr} is the group velocity which is related to the phase velocity v_{ph} by the relationship

$$v_{gr} = \frac{v_0^2}{v_{ph}}, \qquad \text{where } v_0 = (\mu\varepsilon)^{-\frac{1}{2}} \tag{18.162}$$

From Eq. (18.161),

$$\text{Energy stored per unit length} = \left(\frac{1}{v_{gr}}\right) \times \text{Power transmitted} \tag{18.163}$$

$\therefore$ Substituting in (18.160),

$$Q = \frac{\left(\frac{\omega}{v_{gr}}\right)(\text{Power transmitted})}{\text{Power lost per unit length}}$$

$$= \frac{\omega}{2\alpha v_{gr}} \tag{18.164}$$

which can also be written in terms of the cut-off frequency as

$$Q = \left(\frac{\omega v_{ph}}{2\alpha v_0^2}\right) = \frac{\omega}{2\alpha\sqrt{1 - \frac{\omega_c^2}{\omega^2}}} \tag{18.165}$$

Because the waveguides have low attenuation factors compared with the transmission lines, it is possible to construct waveguide sections with extremely high Q's, which are used as resonators or waveguide filters.

It should be noted that the Q factor of ordinary resonant circuits is of the order of a few hundreds, and of resonant lines a few thousands.

18.3.8 Dielectric Slab Waveguide

These are generally used in thin film technology. It consists of two layers of dielectrics. The bottom layer dielectric slab is called the 'substrate'. The top layer is a thin layer of another dielectric of higher dielectric constant (compared to that of the substrate) deposited on the substrate. The propagation loss in the top layer is thus much less at the operating frequency. The slab dielectric waveguide is an asymmetrical structure as shown in Figure 18.21.

Simplifying assumptions: (1) *The substrate is infinitely thick.* This is a justifiable assumption, as the field decays rapidly as we move from the dielectric–substrate interface into the dielectric. (2) The y- and z-dimensions are assumed to be infinite, to make it a simple one-dimensional problem.

As discussed earlier in this chapter (Section 18.3.3), the wave is propagating in the $+z$-direction. Also, either E_z or H_z zero would give rise to modal fields for TE and TM modes respectively. Hence these two independent components must satisfy the wave equation (in rectangular Cartesian coordinates)

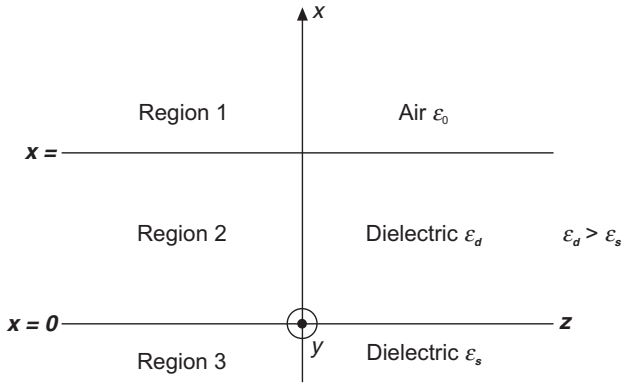


Figure 18.21 Dielectric slab waveguide.

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}+\omega^2\mu\varepsilon\right)(E_z \text{ or } H_z)=0 \tag{18.166}$$

Since the waves are travelling in the z-direction, the z-variation of the fields will be ~ $\exp(-j\beta z)$ where β, the modal phase constant will be determined from the interface continuity conditions. Also, since the structure extends to infinity in the y-direction, the fields are constant functions of y, i.e. independent of y.

$$\therefore \qquad (E_z, H_z) = X(x)\exp(-j\beta z) \tag{18.167}$$

We will now consider the TE mode for this guide.

A. TE Mode

In this case $E_z = 0$, and all the transverse fields are represented in terms of H_z only. Hence,

$$H_z = X(x)\exp(-j\beta z) \tag{18.168}$$

$\exp(j\omega t)$ being implicit in the above expression. Substituting in Eq. (18.166),

$$\frac{1}{X}\frac{d^2X}{dx^2}-\beta^2+\omega^2\mu\varepsilon=0$$

This has to be satisfied for all values of x, and so

$$\frac{1}{X}\frac{d^2X}{dx^2}=\pm k^2 \qquad \text{(say, a constant)}$$

$$\therefore \qquad \frac{d^2X}{dx^2}=\pm k^2X \tag{18.169}$$

The sign of k will have to be chosen so as to satisfy the interface continuity conditions between the media 1 and 2, and 2 and 3. There will be fields in the three media, but in air (1) and in the substrate (3), the fields would decay rapidly as we move away from the interfaces. So this

requirement gives +ve sign of k^2 in the media 1 and 3 and –ve sign in region 2. The equations for the three media can be written as

Region 1 $$\frac{d^2X_1}{dx^2} = k_1^2 X_1$$

$\therefore$ $$k_1^2 - \beta^2 + \omega^2\mu\varepsilon = 0 \tag{18.170a}$$

Region 2 (X_2) $$-k_2^2 - \beta^2 + \omega^2\mu\varepsilon_d = 0 \tag{18.170b}$$

Region 3 (X_3) $$k_3^2 - \beta^2 + \omega^2\mu\varepsilon_s = 0 \tag{18.170c}$$

The field then in the regions 1 and 3 will be in terms of exponentials and in the intermediate region 2 it will be in terms of trigonometric functions. The boundary conditions at infinity would simplify the equations further, i.e. as $x \to +\infty$, $H_{z1} \to 0$ in the region 1 and as $x \to -\infty$, $H_{z3} \to 0$ in region 3. Thus the longitudinal component of the magnetic field in the three regions will be

$$H_{z1} = A_1 \exp(-k_1 x) \exp(-j\beta z) \tag{18.171a}$$

$$H_{z2} = (A_3 \cos k_2 x + A_4 \sin k_2 x) \exp(-j\beta z) \tag{18.171b}$$

$$H_{z3} = A_6 \exp(k_3 x) \exp(-j\beta z) \tag{18.171c}$$

β being same in all the three media.

The transverse field components (H_x, H_y, E_x, E_y) in the three regions can be obtained, as shown in Section 18.3.3, Eqs. (18.59) and (18.62). Substituting for H_z and E_z in these equations, we get:

Region 1

$$E_{x1} = 0, \quad E_{y1} = -\frac{j\omega\mu}{k_1} A_1 \exp(-k_1 x) \exp(-j\beta z) \tag{18.172a}$$

$$H_{x1} = -\frac{j\beta}{k_1} A_1 \exp(-k_1 x) \exp(-j\beta z); \quad H_{y1} = 0$$

Region 2

$$\left.\begin{aligned} &E_{x2} = 0; \quad E_{y2} = \frac{j\omega\mu}{k_z} (A_3 \sin k_2 x - A_4 \cos k_2 x) \exp(-j\beta z) \\ &H_{x2} = \frac{j\beta}{k_2} (A_3 \sin k_2 x - A_4 \cos k_2 x) \exp(-j\beta z) \\ &H_{y2} = 0 \end{aligned}\right\} \tag{18.172b}$$

Region 3

$$\left.\begin{aligned} &E_{x3} = 0; \quad E_{y3} = \frac{j\omega\mu}{k_3} A_6 \exp(k_3 x) \exp(-j\beta z) \\ &H_{x3} = \frac{j\beta}{k_3} A_6 \exp(k_3 x) \exp(-j\beta z) \end{aligned}\right\} \tag{18.172c}$$

The unknowns to be evaluated are A_1, A_3, A_4, A_6, and k_1, k_2, k_3, and β. The interface continuity conditions used for evaluating these unknowns are:

(i) On $x = 0$, $H_{z2} = H_{z3} \rightarrow A_3 = A_6$ (18.173a)

(ii) On $x = 0$, $E_{y2} = E_{y3} \rightarrow \dfrac{-A_4}{k_2} = \dfrac{A_6}{k_3}$ (18.173b)

(iii) On $x = d$, $H_{z1} = H_{z2} \rightarrow A_1 \exp(-k_1 d) = A_3 \cos k_2 d + A_4 \sin k_2 d$ (18.173c)

(iv) On $x = d$, $E_{y1} = E_{y2} \rightarrow \dfrac{-A_1}{k_1} \exp(-k_1 d) = \dfrac{1}{k_2}(A_3 \sin k_2 d - A_4 \cos k_2 d)$ (18.173d)

It should be noted that since there are no surface currents on the dielectric interfaces, the continuity conditions are simplified.

We need three more equations to evaluate k_1, k_2 and k_3, i.e. from Eqs. (18.170a) and (18.170b),

$$k_1^2 + \omega^2 \mu \varepsilon_0 = -k_2^2 + \omega^2 \mu \varepsilon_d$$

or

$$k_1 = \sqrt{[\omega^2 \mu(\varepsilon_d - \varepsilon_s) - k_2^2]} \tag{18.174a}$$

and from Eqs. (18.170b) and (18.170c),

$$k_3 = \sqrt{[\omega^2 \mu(\varepsilon_d - \varepsilon_s) - k_2^2]} \tag{18.174b}$$

Equations (18.173a to d) can be written in matrix form to solve for A_1, A_3, A_4, A_6. The characteristic equation of the determinant of the 4×4 coefficient matrix comes out to be:

$$\tan k_2 d = \frac{(k_1 + k_3)\, k_2}{k_2^2 - k_1 k_3} \tag{18.175}$$

These three equations can be solved as simultaneous equations by a numerical method, and once k_2 has been evaluated, using Eq. (18.170b), i.e.

$$-k_2^2 + \omega^2 \mu \varepsilon_d = \beta^2$$

β can be evaluated.

Some important points which should be noted are:

1. The characteristic Eq. (18.175) has multiple solutions which represent the propagation of 'discrete modes'.
2. k_1, k_2 and k_3 have to be real. So from Eqs. (18.174a) and (18.174b), we have

$$\omega^2 \mu(\varepsilon_d - \varepsilon_0) > k_2^2 \quad \text{and} \quad \omega^2 \mu(\varepsilon_d - \varepsilon_s) > k_2^2 \tag{18.176}$$

   Since k_2^2 is always +ve, this means $\varepsilon_d > \varepsilon_0$ and $\varepsilon_d > \varepsilon_s$. Now $\varepsilon_d > \varepsilon_0$ is always true; and the permittivity of the dielectric slab must be greater than that of the substrate for the propagation of waves to be possible in the slab.
3. For a specified frequency $\omega = 2\pi f$, k_2 must lie between 0 and $\omega\mu(\varepsilon_d - \varepsilon_s)$.

4. When k_2 becomes $> \omega\mu(\varepsilon_d - \varepsilon_s)$, k_3 becomes imaginary and then $\exp(k_3 x)$ does **not** represent an attenuating exponent and now it represents a leakage wave travelling in the $-x$ direction. So this mode is no longer a guided wave mode and is now its cut-off mode. The frequency at which this happens is the cut-off frequency of the guide, i.e.

$$\omega_c = \frac{k_2}{\sqrt{\mu(\varepsilon_d - \varepsilon_s)}} \tag{18.177}$$

18.3.9 Circular (Cylindrical) Dielectric Waveguide

So far, the present discussion has emphasized the mathematics of waveguides with cylindrical geometry. A circular dielectric waveguide is now very commonly used for guiding microwaves as well as millimetre waves. The energy is propagated along the axis of the circular rod made of dielectric material, i.e. glass or silica rod in the form of optical fibres. The operation of the circular waveguide is same as that of dielectric slab metallic waveguide. The modal propagation is a consequence of superposition of totally internally reflected waves inside the dielectric rod. The analysis of this geometry is already explained in Section 18.3.6 and hence without repeating it, the results in terms of field components will be stated now.

A circular dielectric waveguide is shown in Figure 18.21. The rod has a radius a, and permittivity ε_1. To keep the analysis general, let the rod be located in an infinitely large medium of permittivity ε_2. If this is free space, then $\varepsilon_2 = \varepsilon_0$.

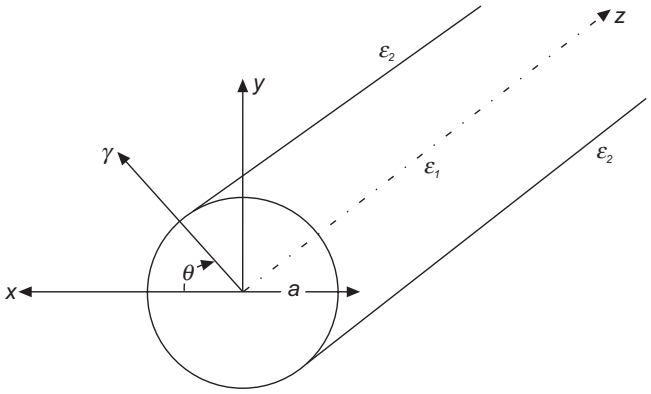


Figure 18.21 Circular dielectric waveguide.

The wave equation for field components will be as derived in Eq. (18.137) and the solutions are given in Eqs. (18.143) to (18.146), So now the TE and TM modes will be considered.

18.3.9.1 Transverse electric mode

For a TE mode, E_z is zero and the longitudinal field H_z is given as

$$H_{z1} = A_1 J_n(k_1 r)\exp(jn\theta)\exp(-j\beta z) \qquad r < a \tag{18.178a}$$

$$H_{z2} = G_1 K_n(k_2 r)\exp(jn\theta)\exp(-j\beta z) \qquad r > a \tag{18.178b}$$

The transverse fields are obtained as:

For $r < a$ (i.e. inside the rod)

$$E_{r1} = \frac{-j\omega\mu}{k_1^2}\frac{1}{r}\frac{\partial H_{z1}}{\partial\theta} = A_1\frac{\omega\mu}{k_1^2}\frac{n}{r}J_n(k_1 r)\cdot\exp(jn\theta)\cdot\exp(-j\beta z) \tag{18.179a}$$

$$E_{\theta 1} = \frac{j\omega\mu}{k_1^2}\frac{\partial H_{z1}}{\partial r} = A_1\frac{j\omega\mu}{k_1}J_n'(k_1 r)\exp(jn\theta)\cdot\exp(-j\beta z) \tag{18.179b}$$

$$H_{r1} = -\frac{j\beta}{k_1^2}\frac{\partial H_{z1}}{\partial r} = A_1\frac{-j\beta}{k_1}J_n'(k_1 r)\exp(jn\theta)\cdot\exp(-j\beta z) \tag{18.179c}$$

$$H_{\theta 1} = -\frac{j\beta}{k_1^2}\frac{1}{r}\frac{\partial H_z}{\partial\theta} = A_1\frac{\beta}{k_1^2}\frac{n}{r}J_n(k_1 r)\exp(jn\theta)\cdot\exp(-j\beta z) \tag{18.179d}$$

and for $r > a$ (i.e. outside the waveguide)

$$E_{r2} = \frac{-j\omega\mu}{-k_2^2}\frac{1}{r}\frac{\partial H_{z2}}{\partial\theta} = G_1\frac{-\omega\mu}{k_2^2}\frac{n}{r}K_n(k_2 r)\exp(jn\theta)\exp(-j\beta z) \tag{18.180a}$$

$$E_{\theta 2} = \frac{j\omega\mu}{-k_2^2}\frac{\partial H_{z2}}{\partial r} = G_1\frac{-j\omega\mu}{k_2}K_n'(k_2 r)\exp(jn\theta)\exp(-j\beta z) \tag{18.180b}$$

$$H_{r2} = \frac{-j\beta}{-k_1^2}\frac{\partial H_{z2}}{\partial r} = G_1\frac{j\beta}{k_2}K_n'(k_2 r)\exp(jn\theta)\exp(-j\beta z) \tag{18.180c}$$

$$H_{\theta 2} = \frac{-j\beta}{-k_2^2}\frac{1}{r}\frac{\partial H_{z2}}{\partial\theta} = G_1\frac{-\beta}{k_2^2}\frac{n}{r}K_n(k_2 r)\exp(jn\theta)\exp(-j\beta z) \tag{18.180d}$$

To evaluate the constants of integration, the relevant boundary conditions are that the tangential components of **E** and **H** are continuous on the surface of the rod (i.e. at $r = a$). This gives

$$\left.\begin{aligned} &H_{z1} = H_{z2} \quad \text{and} \quad E_{\theta 1} = E_{\theta 2} \text{ at } r = a \\ \text{i.e.} \quad &A_1 J_n(ka) = G_1 K_n(k_1 a) \quad \text{and} \quad \frac{A_1}{k_1}J_n'(ka) = -\frac{G_1}{k_2}K_n'(k_1 a) \end{aligned}\right\} \tag{18.181}$$

Since H_θ is also tangential to the rod surface, we get

$$H_{\theta 1} = H_{\theta 2} \qquad \text{or} \qquad -\frac{nA_1}{k_1^2}J_n(ka) = \frac{nG_1}{k_2^2}K_n(k_1 a) \tag{18.182}$$

The above equations for H_z and H_θ can be simultaneously satisfied only if β or G_1 or n is made identically equal to zero.

Now G_1 which represents the field amplitude cannot be made zero. Also β cannot be zero because this is a travelling mode.

$$\therefore \qquad n = 0 \text{ is the required condition} \tag{18.183}$$

This means that there is no θ variation, i.e. the field is circularly symmetric.

$$E_{r1} = E_{r2} = 0 \qquad \text{and} \qquad H_{\theta 1} = H_{\theta 2} = 0 \tag{18.184}$$

Hence the resulting field components are:
Inside the waveguide ($r < a$),

$$H_{z1} = A_1 J_0(k_1 r) \exp(-j\beta z) \tag{18.185a}$$

$$E_{\theta 1} = A_1 \frac{j\omega\mu}{k_1} J_0'(k_1 r) \exp(-j\beta z) \tag{18.185b}$$

$$H_{r1} = -A_1 \frac{j\beta}{k_1} J_0'(k_1 r) \exp(-j\beta z) \tag{18.185c}$$

Outside the waveguide ($r > a$),

$$H_{z2} = G_1 K_0(k_2 r) \exp(-j\beta z) \tag{18.186a}$$

$$E_{\theta 2} = -G_1 \frac{j\omega\mu}{k_2} K_0'(k_2 r) \exp(-j\beta z) \tag{18.186b}$$

$$H_{r2} = G_1 \frac{j\beta}{k_2} K_0'(k_2 r) \exp(-j\beta z) \tag{18.186c}$$

To eliminate the arbitrary constants, the characteristic equation for the TE mode is obtained from

$$\frac{E_{\theta 1}}{H_{z1}} = \frac{E_{\theta 2}}{H_{z2}} \text{ on } r = a,$$

i.e.

$$\frac{J_0'(k_1 a)}{k J_0(k_1 a)} = \frac{K_0'(k_2 a)}{k_1 K_0(k_2 a)} \tag{18.187}$$

This equation has multiple roots, because J_0 is an orthogonal function. A TE mode is designated as TE_{0m} because the suffix 0 corresponds to $n = 0$.

18.3.9.2 Transverse magnetic mode (TM)

For the transverse magnetic mode $H_z = 0$, and the transverse fields are expressed in terms of E_z. Following the analysis of the TE mode, it will be seen that the TM mode is also circularly symmetric and hence the corresponding field components are:

$$E_{z1} = A_1 J_0(k_1 r) \exp(-j\beta z) \tag{18.188a}$$

$$E_{z2} = G_1 K_0(k_2 r) \exp(-j\beta z) \tag{18.188b}$$

The corresponding transverse components are:
Inside the waveguide ($r < a$),

$$E_{r1} = \frac{-j\beta}{k_1^2} \frac{\partial E_{z1}}{\partial r} = \frac{-j\beta}{k_1} A_1 J_0'(k_1 r) \exp(-j\beta z) \tag{18.189a}$$

$$H_{\theta 1} = \frac{-j\omega\varepsilon_1}{k_1^2}\frac{\partial E_{z1}}{\partial r} = \frac{-j\omega\varepsilon_1}{k_1} A_1 J_0'(k_1 r)\exp(-j\beta z) \tag{18.189b}$$

and outside the waveguide ($r > a$),

$$E_{r2} = \frac{-j\beta}{-k_2^2}\frac{\partial E_{z2}}{\partial r} = \frac{j\beta}{k_2} G_1 K_0'(k_2 r)\exp(-j\beta z) \tag{18.190a}$$

$$H_{\theta 2} = \frac{-j\omega\varepsilon_2}{-k_2^2}\frac{\partial E_{z2}}{\partial r} = \frac{j\omega\varepsilon_2}{k_2} G_1 K_0'(k_2 r)\exp(-j\beta z) \tag{18.190b}$$

Next, the continuity of the tangential components of the fields at the waveguide surface $r = a$ gives

$$E_{z1} = E_{z2} \quad \text{and} \quad H_{\theta 1} = H_{\theta 2} \tag{18.191}$$

from which we get:

$$A_1 J_0(k_1 a) = G_1 K_0(k_2 a) \quad \text{and} \quad -\frac{\varepsilon_1}{k_1} A_1 J_0'(k_1 a) = \frac{\varepsilon_2}{k_2} G_1 K_0'(k_2 a) \tag{18.192}$$

Taking the ratios of these two equations,

$$\frac{\varepsilon_1}{k_1}\frac{J_0'(k_1 a)}{J_0(k_1 a)} = -\frac{\varepsilon_2}{k_2}\frac{K_0'(k_2 a)}{K_0(k_2 a)} \tag{18.193}$$

This is the characteristic equation of the TM mode, which is similar to that of the TE mode. The above equation also has multiple roots. The TM mode is designated as TM_{0m} mode.

18.3.9.3 Cut-off frequencies of TM mode

For the two Bessel functions, i.e. $J_0(k_1 r)$ the Bessel function for the region inside the dielectric rod and $K_0(k_2 r)$ the modified Bessel function for the region outside the rod, K_0 is not oscillatory and is a decaying function as shown in Figure 18.16(b). We also have

$$k_1 = \sqrt{(\omega^2\mu\varepsilon_1 - \beta^2)} \quad \text{and} \quad jk_2 = j\sqrt{(\beta^2 - \omega^2\mu\varepsilon_2)} \tag{18.194a}$$

$$\therefore \quad k_1^2 + k_2^2 = \omega^2\mu(\varepsilon_1 - \varepsilon_2) = \text{constant} \tag{18.194b}$$

for a given waveguide and frequency.

The decaying field outside the rod remains so, so long as k_2 is real. If k_2 were to be imaginary K_0 would be replaced by ordinary Bessel function which is oscillatory in nature and travelling radially away from the axis of the rod—a leaking wave. This then no longer remains a guided mode. Hence the cut-off frequency would be obtained when k_2 is made zero. From Eq. (18.194b), it is obvious that both k_1 and k_2 cannot be made simultaneously equal to zero.

Hence, as $k_2 \to 0$, since k_1 cannot approach zero, $J_0(k_1 a) \to 0$

$\therefore$ $J_0(k_1 a) = 0$ corresponds to the cut-off frequency of TE or TM mode, i.e.

$$k_c a = \chi_{0m} \tag{18.195a}$$

where χ_{0m} is the mth root of the J_0 Bessel function.

∴ The modal cut-off frequency

$$\omega_c = \frac{\chi_{0m}}{a\sqrt{\mu(\varepsilon_1 - \varepsilon_2)}} \tag{18.195b}$$

The first root of the J_0 function is 2.40, and hence the lowest frequency which can propagate in TE or TM mode is

$$\omega_{c01} = \frac{2.4}{a\sqrt{\mu(\varepsilon_1 - \varepsilon_2)}} \tag{18.195c}$$

18.3.9.4 Hybrid modes

So far it has been seen that in circular dielectric waveguides the TE and TM modes are circularly symmetric, This is so when either $E_z = 0$ or $H_z = 0$, i.e. one field component in the direction of propagation is mode equal to zero. When both E_z and H_z are non-zero, then the mode is neither TE nor TM. Such fields are called 'hybrid modal fields.' When such fields exist, they are not circularly symmetric. The longitudinal fields in the two media can then be written as follows:

For $r < a$, i.e. inside the waveguide

$$E_{z1} = A\, J_n(k_1 r) \exp(jn\theta - j\beta z) \tag{18.196a}$$

$$H_{z1} = B\, J_n(k_1 r) \exp(jn\theta - j\beta z) \tag{18.196b}$$

For $r > a$, i.e. outside the waveguide

$$E_{z2} = C\, K_n(k_2 r) \exp(jn\theta - j\beta z) \tag{18.196c}$$

$$H_{z2} = D\, K_n(k_2 r) \exp(jn\theta - j\beta z) \tag{18.196d}$$

A, B, C, D being the amplitudes of the respective field components.

From these four equations for the z-components of the field, the transverse field expressions can be written down as follows:

Inside the waveguide, $r < a$

$$E_{r1} = \frac{\omega\mu n}{k_1^2 r} B\, J_n(k_1 r) \exp(jn\theta) - \frac{j\beta}{k_1} A\, J_n'(k_1 r) \exp(jn\theta) \tag{18.197a}$$

$$E_{\theta 1} = \frac{j\omega n}{k_1} B\, J_n'(k_1 r) \exp(jn\theta) + \frac{\beta}{k_1^2} \frac{nA}{r} J_n(k_1 r) \exp(jn\theta) \tag{18.197b}$$

$$H_{r1} = -\frac{\omega\varepsilon_1}{k_1^2} \frac{n}{r} A\, J_n(k_1 r) \exp(jn\theta) - \frac{j\beta}{k_1} BJ_n'(k_1 r) \exp(jn\theta) \tag{18.197c}$$

$$H_{\theta 1} = \frac{-j\omega\varepsilon_1}{k_1} A\, J_n'(k_1 r) \exp(jn\theta) + \frac{\beta n}{k_1^2 r} BJ_n(k_1 r) \exp(jn\theta) \tag{18.197d}$$

Outside the waveguide, $r > a$,

$$E_{r2} = -\frac{\omega\mu}{k_2^2}\frac{n}{r} D\, K_n(k_2 r)\exp(jn\theta) + \frac{j\beta}{k_2} C\, K_n'(k_2 r)\exp(jn\theta) \tag{18.198a}$$

$$E_{\theta 2} = -\frac{j\omega\mu}{k_2} D\, K_n'(k_2 r)\exp(jn\theta) - \frac{\beta}{k_2^2}\frac{n}{r} C\, K_n(k_2 r)\exp(n\theta) \tag{18.198b}$$

$$H_{r2} = \frac{\omega\varepsilon_2}{k_2^2}\frac{n}{r} C\, K_n(k_2 r)\exp(jn\theta) + \frac{j\beta}{k_2} D\, K_n'(k_2 r)\exp(jn\theta) \tag{18.198c}$$

$$H_{\theta 2} = \frac{j\omega\varepsilon_2}{k} C\, K_n'(k_2 r)\exp(jn\theta) - \frac{\beta}{k_2^2}\frac{n}{r} K_n(k_2 r)\exp(jn\theta) \tag{18.198d}$$

For four unknowns A, B, C, D to be evaluated, the four boundary conditions are:

at $r = a$,
$$E_{z1} = E_{z2}, \qquad H_{z1} = H_{z2}$$
$$E_{\theta 1} = E_{\theta 2}, \qquad H_{\theta 1} = H_{\theta 2}$$

The four equations obtained from these four conditions can be written in a compact matrix form as:

$$\begin{bmatrix} J_n(k_1a) & 0 & -K_n(k_2a) & 0 \\ 0 & J_n(k_1a) & 0 & -K_n(k_2a) \\ \dfrac{\beta n}{ak_1^2}J_n(k_1a) & \dfrac{j\omega\mu}{k_1}J_n'(k_1a) & \dfrac{\beta n}{ak_2^2}K_n(k_2a) & \dfrac{j\omega\mu}{k_2}K_n'(k_2a) \\ \dfrac{-j\omega\varepsilon_1}{k_1}J_n'(k_1a) & \dfrac{\beta n}{ak_1^2}J_n(k_1a) & \dfrac{-j\omega\varepsilon_2}{k_2}K_n'(k_2a) & \dfrac{\beta n}{ak_2^2}K_n(k_2a) \end{bmatrix} \begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix} = 0 \tag{18.199}$$

The characteristic equation of the hybrid mode can be obtained by setting the determinant of the coefficient matrix of the last equation to zero. If then, in the characteristic equation n is made equal to zero, the equation splits into two equations which are

$$\frac{J_0'(k_1a)}{k_1 J_0(k_1a)} + \frac{K_0'(k_2a)}{k_2 K_0(k_2a)} = 0 \tag{18.200a}$$

and

$$\frac{\varepsilon_1}{k_1}\frac{J_0'(k_1a)}{J_0(k_1a)} + \frac{\varepsilon_2}{k_2}\frac{K_0'(k_2a)}{K_0(k_2a)} = 0 \tag{18.200b}$$

Equations (18.200a) and (18.200b) are the characteristic equations of TE and TM modes respectively [i.e. Eqs. (18.187) and (18.193)].

18.3.9.5 Cut-off frequency of a hybrid mode

Evaluation of the cut-off frequency of a hybrid mode is rather complicated. The recurrence relations of Bessel functions have to be used. The derivatives of Bessel functions are replaced by Bessel functions of higher and lower orders. After a somewhat lengthy algebraic manipulation, it is

found that the lowest mode is HE_{11} and for $n = 1$, its cut-off is given by $k_{1c} = 0$. The cut-off frequency for $k_{1c} = 0$ is 0. Hence it means that HE_{11} does not have a cut-off frequency, i.e. it can propagate any frequency up to dc. It is an interesting point to be noted that for direct currents, the fields inside a dielectric rod are hybrid and **not** transverse. So the electrostatic field (which corresponds to dc) must have a field component along the axis of the dielectric rod.

18.4 RESONANT CAVITIES (CAVITY RESONATORS)

A tuned circuit consisting of a coil and a capacitor (i.e. a resonating circuit) acts as an energy storage device, and is an important part of a radio receiver. But at very high frequencies, a simple coil cannot act as an inductor. In the centimetre range of wavelengths (i.e. 100 MHz and above), the dimensions of the circuits are comparable with the operating wavelength and hence unwanted radiation takes place. So, at high frequencies the RLC circuits are replaced by cavity resonators. Such resonant cavities are used in klystron tubes, band-pass filters and wave-meters. Even the microwave oven is an application of a cavity resonator so that the oven is the cavity itself which is supplied by a waveguide feed and a power supply.

In a resonant cavity, the electromagnetic waves are reflected by the conducting walls and thus the resonance is produced in it. A cavity resonator can be of any shape—rectangular, cylindrical or spherical—but an important class is produced by placing end-faces on a finite length (i.e. piece) of cylindrical waveguide.

We will consider a rectangular cavity (or a closed conducting box). It is effectively a rectangular waveguide shorted at both the ends. Now when a metal plate is placed at the end of a waveguide (rectangular), the propagation of the electromagnetic wave is blocked and the wave is reflected. So now there are two waves travelling in opposite directions. A standing wave pattern is produced which must satisfy the boundary conditions at both the ends. The problem can be analysed either (1) by starting with the (say) longitudinal magnetic wave travelling in the direction of propagation for TE_{mn} mode or electric wave for the TM_{mn} mode, or (2) by solving the wave equation for the closed box with appropriate boundary conditions.

18.4.1 TE Mode Waveguide Approach

In the waveguide, the wave is assumed to travel in the z-direction, and for the cavity, the metal plates are placed at $z = 0$ and $z = d$. In this situation, the two waves are travelling in opposite directions (i.e. $+z$ and $-z$). For the TE_{mn} mode, the longitudinal H_z for the travelling wave is given by

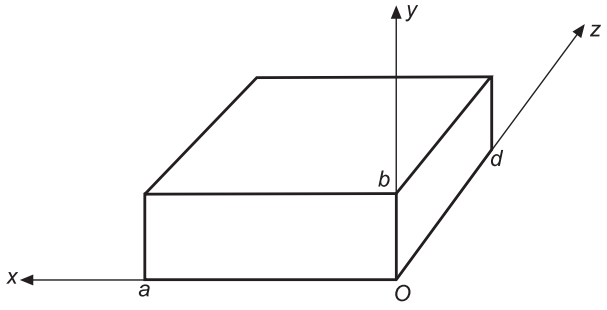


Figure 18.22 A rectangular resonant cavity.

$$H_z = H_0 \cos\frac{m\pi x}{a}\cos\frac{n\pi y}{b}\exp[j(\omega t - \beta z)] \tag{18.201a}$$

For the reflected TE_{mn} wave, let its amplitude be H_0'. Hence

$$H_z' = H_0' \cos\frac{m\pi x}{a}\cos\frac{n\pi y}{b}\exp[j(\omega t + \beta z)] \tag{18.201b}$$

The total H field inside the cavity is then

$$H_Z = H_z + H_z' \tag{18.202}$$

Since H_Z would be perpendicular to the boundaries $z = 0$ and $z = d$, it follows that it should be zero on these two planes.

$$\therefore \qquad \text{at } z = 0, \quad H_z = 0 \quad \rightarrow \quad H_0 = -H_0' \tag{18.203}$$

$$H_z = H_0 \cos\frac{m\pi x}{a}\cos\frac{n\pi y}{b}[\exp(-j\beta z) - \exp(j\beta z)]$$

$$= -2jH_0 \cos\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi y}{b}\right)\sin(\beta z) \tag{18.204}$$

not writing $\exp(j\omega t)$ in these expressions explicitly.

Also on $z = d$, $H_z = 0 \quad \rightarrow \quad \beta d = p\pi$, $p = 1, 2, 3, \ldots$

or

$$\beta = \frac{p\pi}{d} \tag{18.205}$$

Substituting for β from this equation, into Eqs. (18.67) or (18.70), while noting that $\gamma = j\beta$ in the present case,

$$\omega^2\mu\varepsilon = \left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 + \left(\frac{p\pi}{d}\right)^2$$

or

$$\omega = 2\pi f = \frac{1}{\sqrt{\mu\varepsilon}}\left[\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 + \left(\frac{p\pi}{d}\right)^2\right]^{1/2} \tag{18.206}$$

where m, n, p are all integers and hence there will be discrete frequencies at which the electromagnetic fields can be excited inside the cavity. Since the fields are characterized by three indices, this mode is designated as TE_{mnp}.

18.4.2 TE Mode—From the Fundamentals

It should be noted that depending on how the cavity is excited, the wave can propagate in any of the x-, y-, or z- direction. For simplicity, we choose the direction of propagation as $+z$ direction (as in Section 18.4.1, Figure 18.22).

Since

$$E_z = 0, \quad H_z = X(x)\,Y(y)\,Z(z) \tag{18.207}$$

where H_z is the solution of $\nabla^2\mathbf{H} + k_2\mathbf{H} = 0$.

$$\left.\begin{aligned} X(x) &= B_1 \cos k_x x + B_2 \sin k_x x \\ Y(y) &= B_3 \cos k_y y + B_4 \sin k_y y \\ Z(z) &= B_5 \cos k_z z + B_6 \sin k_z z \end{aligned}\right\} \tag{18.208}$$

where
$$k^2 = k_x^2 + k_y^2 + k_z^2 \tag{18.209}$$

The boundary conditions are:

(i) $H_z = 0$ at $z = 0, d$

(ii) $\dfrac{\partial H_z}{\partial x} = 0$ at $x = 0, a$

(iii) $\dfrac{\partial H_z}{\partial y} = 0$ at $y = 0, b.$

From the boundary conditions (ii) and (iii),

$$B_2 = 0, \quad B_4 = 0, \quad \text{and} \quad k_x = \frac{m\pi}{a}, \quad k_y = \frac{n\pi}{b} \tag{18.210}$$

where m = 0, 1, 2, 3, ... and n = 0, 1, 2, 3.

For the boundary condition (i),

$$B_5 = 0 \quad \text{and} \quad k_z = \frac{p\pi}{d}, \quad \text{where } p = 1, 2, 3, ... \tag{18.211}$$

$$\therefore \quad H_z = H_0 \cos\left(\frac{m\pi x}{a}\right) \cos\left(\frac{n\pi y}{b}\right) \sin\left(\frac{p\pi z}{d}\right) \tag{18.212}$$

$$\beta^2 = k^2 = \left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 + \left(\frac{p\pi}{d}\right)^2 = \omega^2 \mu\varepsilon$$

$$\therefore \quad \omega = 2\pi f = \frac{1}{\sqrt{\mu\varepsilon}} \left[\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 + \left(\frac{p\pi}{d}\right)^2\right]^{1/2} \tag{18.213}$$

The resonant frequency is same as derived by the other method, i.e. Eq. (18.183). It should be noted that both m and n cannot be zero simultaneously because the field components will be zero if both m and n are zeroes. The 'dominant mode' (i.e. the mode having the lowest resonant frequency is TE_{101} mode ($a > b < c$). A practical resonant cavity has walls with finite conductivity σ_c, and hence can lose stored energy. The loss is determined by the quality factor Q.

18.4.3 TM Mode—From the Fundamentals

In this case, $H_z = 0$ and hence

$$E_z = X(x)\ Y(y)\ Z(z) \tag{18.214}$$

$$\therefore \qquad \left.\begin{aligned} X(x) &= A_1 \cos k_x x + A_2 \sin k_x x \\ Y(y) &= A_3 \cos k_y y + A_4 \sin k_y y \\ Z(z) &= A_5 \cos k_z z + A_6 \sin k_z z \end{aligned}\right\} \tag{18.215}$$

where $$k^2 = k_x^2 + k_y^2 + k_z^2 = \omega^2 \mu\varepsilon \tag{18.216}$$

The boundary conditions are:

$$\begin{aligned} &\text{(i)} \quad E_z = 0 \quad \text{at} \quad x = 0, a \\ &\text{(ii)} \quad E_z = 0 \quad \text{at} \quad y = 0, b \\ &\text{(iii)} \quad E_x = 0, \quad E_y = 0 \quad \text{at} \quad z = 0, d \end{aligned} \tag{18.217}$$

From the boundary conditions (i) and (ii),

$$\left.\begin{aligned} &A_1 = 0, \quad A_3 = 0, \quad k_x = \frac{m\pi}{a}, \quad k_y = \frac{n\pi}{b} \\ &\text{where} \qquad m = 1, 2, 3, \ldots \qquad n = 1, 2, 3, \ldots \end{aligned}\right\} \tag{18.218}$$

Next, for the boundary condition (iii), from the two curl equations of Maxwell, i.e.

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} \quad \text{and} \quad \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{18.218a}$$

we get

$$j\omega\varepsilon E_x = \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} \quad \text{and} \quad -j\omega\mu H_y = \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x}$$

Combining these two equations,

$$j\omega\varepsilon E_x = \frac{\partial H_z}{\partial y} + \frac{1}{j\omega\mu}\left(\frac{\partial^2 E_x}{\partial z^2} - \frac{\partial^2 E_z}{\partial x \partial z}\right)$$

Since $H_z = 0$, this simplifies to

$$j\omega\varepsilon E_x = \frac{1}{j\omega\mu}\left(\frac{\partial^2 E_x}{\partial z^2} - \frac{\partial^2 E_z}{\partial z \partial x}\right) \tag{18.219}$$

and considering E_y and H_x from the same two $(\nabla\times)$ equations

$$j\omega\varepsilon E_y = \frac{1}{-j\omega\mu}\left(\frac{\partial^2 E_z}{\partial y \partial z} - \frac{\partial^2 E_y}{\partial z^2}\right) \tag{18.220}$$

From these two equations [i.e. (18.219) and (18.220)], to satisfy the boundary condition (iii) which states

$$E_x = 0, \quad E_y = 0 \quad \text{at} \quad z = 0, d$$

the requisite condition is

$$\frac{\partial E_z}{\partial z} = 0 \quad \text{at} \quad z = 0 \text{ and } d.$$

$$\rightarrow A_6 = 0 \quad \text{and} \quad \sin k_z d = 0 = \sin p\pi$$

$$\therefore \qquad k_z = \frac{p\pi}{d}, \quad \text{where } p = 0, 1, 2, 3, ... \tag{18.221}$$

Substituting for the unknowns in Eqs. (18.214) and (18.215),

$$E_z = E_0 \sin\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right) \cos\left(\frac{p\pi z}{d}\right) \tag{18.222}$$

where $E_0 = A_2 A_4 A_5$.

The other field components can be obtained from Eq. (18.222) and the two curl Eqs. (18.218a).

The phase constant comes out to be

$$k^2 = \left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 + \left(\frac{p\pi}{d}\right)^2$$

$$= \omega^2 \mu\varepsilon$$

$$\therefore \qquad \omega = 2\pi f = \frac{1}{\sqrt{\mu\varepsilon}} \sqrt{\left[\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 + \left(\frac{p\pi}{d}\right)^2\right]} \tag{18.223}$$

The lowest order TM mode is TM_{110}

18.4.4 Stored Energy and Quality Factor

Considering the TE_{101} mode (m = 1, n = 0), the longitudinal magnetic field component is [from Eq. (18.204)]

$$H_z = -2j\, H_0 \cos\left(\frac{\pi x}{a}\right) \sin\left(\frac{\pi z}{d}\right) \tag{18.224}$$

The other field components come out to be

$$H_x = 2j \frac{a}{d} H_0 \sin\left(\frac{\pi x}{a}\right) \cos\left(\frac{\pi z}{d}\right) \tag{18.225}$$

$$E_y = -2\omega_0 \mu H_0 \frac{a}{\pi} \sin\left(\frac{\pi x}{a}\right) \sin\left(\frac{\pi z}{d}\right) \tag{18.226}$$

by suitable substitutions in Eqs. (18.77c) and (18.77b) respectively (for the incident and the reflected waves in the cavity).

The resonant frequency ω_0 for the TE_{101} mode comes out as

$$\omega_0 = 2\pi f_0 = \frac{1}{\sqrt{\mu\varepsilon}}\left[\left(\frac{\pi}{a}\right)^2 + \left(\frac{\pi}{d}\right)^2\right]^{1/2} \tag{18.227}$$

Since every resonant circuit stores electromagnetic energy, we now calculate the stored energy.

The average electrical energy stored inside the cavity is

$$W_e = \frac{\varepsilon}{2}\int_{x=0}^{a}\int_{y=0}^{b}\int_{z=0}^{d}\frac{1}{2}\mathrm{Re}\,(\mathbf{E}.\mathbf{E}^*)\,dx\,dy\,dz$$

$$= \varepsilon|H_0|^2\left(\frac{\omega\mu a}{\pi}\right)^2\iiint\sin^2\left(\frac{\pi x}{a}\right)\sin^2\left(\frac{\pi z}{d}\right)dx\,dy\,dz$$

$$= \frac{a\,b\,d\,\varepsilon}{4}|H_0|^2\left(\frac{\omega\mu a}{\pi}\right)^2 \tag{18.228}$$

The average magnetic energy stored inside the cavity

$$W_m = \frac{\mu}{2}\int_{x=0}^{a}\int_{y=0}^{b}\int_{z=0}^{d}\frac{1}{2}\mathrm{Re}\,(\mathbf{H}\cdot\mathbf{H}^*)\,dx\,dy\,dz$$

$$= \mu|H_0|^2\int_{x=0}^{a}\int_{y=0}^{b}\int_{z=0}^{d}\left[\left(\frac{a}{d}\right)^2\sin^2\left(\frac{\pi x}{a}\right)\cos^2\left(\frac{\pi z}{d}\right) + \cos^2\left(\frac{\pi x}{a}\right)\sin^2\left(\frac{\pi z}{d}\right)\right]dx\,dy\,dz$$

$$= \frac{abd\mu}{4}|H_0|^2\left[\left(\frac{a}{d}\right)^2 + 1\right] = W_e \qquad \left(\text{since } \left(\frac{a}{d}\right)^2 + 1 = \omega^2\mu\varepsilon a\pi\right) \tag{18.229}$$

which is obtained from Eq. (18.223) or (18.213) by suitable substitutions for m, n, p.

Thus the average stored electrical and magnetic energy are equal for the cavity resonator. This equality is identical with the LC circuit at the resonance.

The quality factor Q, for the resonator, is a measure of loss in the resonator, and is defined as

$$Q = \omega_0\,\frac{\text{Energy stored in the cavity}}{\text{Power loss in the cavity}} \tag{18.230}$$

$$= \omega_0\,\frac{W_e + W_m}{\text{Power loss in the cavity}}$$

$$= \frac{2\omega_0 W_e}{\text{Power loss in the cavity}} \tag{18.231}$$

Assuming the loss in the cavity to be small, the field distribution of a lossy cavity is almost same as that of a loss-less cavity. The surface currents on the walls of the waveguide (and the conductor loss) can be obtained from the knowledge of the magnetic field. If the surface resistance is denoted by R_S,

$$R_S = \sqrt{(\omega_0 \mu)/(2\sigma)},\ \sigma = \text{conductivity of the cavity walls,}$$

then for the TE_{101} mode,

$$Q_{101} = \frac{\mu^2 \varepsilon (\omega_0 ad)^3 b}{2\pi^2 R_S [2b(a^3 + d^3) + ad\,(a^2 + d^2)]} \tag{18.232a}$$

$$= \frac{(a^2 + d^2)abd}{\delta[2b(a^3 + d^3) + ad(a^2 + d^2)]} \tag{18.232b}$$

where $\delta = \sqrt{\dfrac{2}{\omega_0 \mu_0 \sigma}}$ = depth of penetration (or skin depth) = $d\sqrt{2}$ of Eq. (15.7).

Note: '*d*' of skin-depth or depth of penetration $(d\sqrt{2})$ of Eq. (15.7) is not to be confused with the dimension *d* of the cavity. Hence the notation δ has been used here.

18.4.5 Cylindrical Cavity

It is a cylindrical piece of radius *R* and axial length *d*, and its walls are assumed to have infinite conductivity. For simplicity, the cavity is asumed to be filled with a loss-less dielectric of characteristics μ, ε. Since the reflections take place on the end-surfaces of the cylinder the relevant expressions for the standing waves (using the cylindrical coordinate system) are:

$$A\sin(kz) + B\cos(kz)$$

Since the two ends of the cylindrical cavity are $z = 0$ and $z = d$, the boundary conditions can be satisfied only if

$$k = \frac{p\pi}{d}, \qquad \text{where } p = 0, 1, 2, 3, \ldots \tag{18.233}$$

For TE modes, the vanishing of H_z at $z = 0$ and $z = d$ requires

$$H_z = \psi(x, y) \sin\left(\frac{p\pi z}{d}\right), \qquad p = 1, 2, 3, \ldots \tag{18.234}$$

For TM modes, the vanishing of E_t at $z = 0$ and $z = d$ requires

$$E_z = \psi(x, y) \cos\left(\frac{p\pi z}{d}\right), \qquad p = 0, 1, 2, 3, \ldots \tag{18.235}$$

where ψ is a scalar function which satisfies the two-dimensional wave equation

$$(\nabla_t^2 + \gamma^2)\psi = 0 \tag{18.236}$$

and ∇_t^2 is the transverse part of the Laplacian operator, i.e.

$$\nabla_t^2 \equiv \nabla^2 - \frac{\partial^2}{\partial z^2} \tag{18.237}$$

In the Cartesian coordinates, the transverse fields for the two modes are obtained as:

TM mode:
$$E_t = -\left[\left(\frac{p\pi}{\gamma^2 d}\right)\sin\left(\frac{p\pi z}{d}\right)\right]\nabla_t\psi \tag{18.238a}$$

$$H_t = \left[\left(\frac{j\varepsilon\omega}{\gamma^2 \mathrm{c}}\right)\cos\left(\frac{p\pi z}{d}\right)\right]\mathbf{i}_z \times \nabla_t\psi \tag{18.238b}$$

TE mode:
$$E_t = -\left[\left(\frac{j\omega\mu}{\gamma^2 \mathrm{c}}\right)\sin\left(\frac{p\pi z}{d}\right)\right]\mathbf{i}_z \times \nabla_t\psi \tag{18.239a}$$

$$H_t = \left[\left(\frac{p\pi}{\gamma^2 d}\right)\cos\left(\frac{p\pi z}{d}\right)\right]\nabla_t\psi \tag{18.239b}$$

where the constant
$$\gamma^2 = \left(\frac{\omega^2\mu\varepsilon}{\mathrm{c}}\right) - \left(\frac{p\pi}{d}\right)^2 \tag{18.240}$$

and ψ can be solved as an eigenvalue problem.

In the cylindrical resonant cavity, the tuning is done with a piston by varying d. For the TM mode, the transverse wave equation for ψ is = E_z, subject to the boundary condition that $E_z = 0$ at $r = R$.

The solution for the cavity is

$$\psi(r, \phi) = J_m(\gamma_{mn} r)\exp(\pm jm\phi)$$
$$\text{where} \qquad \gamma_{mn} = \frac{x_{mn}}{R} \tag{18.241}$$

x_{mn}, being the nth root of $J_m(x) = 0$.

The integers m and n have the values

$$m = 0, 1, 2, 3, \ldots; \qquad n = 1, 2, 3, \ldots$$

The resonant frequencies are given by:

$$\omega_{mnp} = \frac{\mathrm{c}}{\sqrt{\mu\varepsilon}}\sqrt{\left(\frac{x_{mn}^2}{R^2} + \frac{p^2\pi^2}{d^2}\right)} \tag{18.242}$$

For the lowest TM mode, $m = 0$, $n = 1$, $p = 0$, the designation is TM_{010} and its resonance frequency is

$$\omega_{010} = \frac{2.405\mathrm{c}}{R\sqrt{\mu\varepsilon}} \tag{18.243}$$

In general terms, for cavities of a few cm linear dimension with α of about 10^{-4} cm, the Q value is of the order of 10,000

PROBLEMS

18.1 Show that the input impedance of a loaded lossy transmission line is given by

$$Z_{in} = Z_c \left[\frac{Z_L + jZ_c \tanh \gamma l}{Z_c + jZ_l \tanh \gamma l} \right]$$

where

γ = propagation constant
Z_c = characteristic impedance
Z_L = load impedance
l = length of the line.

Hence, for a quarter wavelength line

$$Z_{in} = \frac{Z_c^2}{Z_L}$$

18.2 A coil has a complex impedance of resistance R and self-inductance L. It is connected in parallel with a capacitor of capacitance C with an imperfect dielectric equivalent to a series resistance which is also R. Find (i) a value of R which makes the impedance purely resistive at all values of ω, and (ii) a value of ω which again makes the impedance purely resistive for all R.

Ans.: (i) $R = \sqrt{\frac{L}{C}}$, (ii) $\omega = \frac{1}{\sqrt{LC}}$

18.3 Rectangular waveguide is often made of brass or steel for economy, and then silver-plated to provide the lowest losses. Assuming operation at 10 GHz with $\sigma = 6.17 \times 10^7$ mho/m for silver, calculate the amount of silver required per mile to provide a three-skin-depths coating on a waveguide with an inner periphery of 10 cm. Density of silver = 10.5 g/cc.

Ans.: 3.25 kg

18.4 Define phase velocity v_{ph} and group velocity v_{gr} of a travelling wave, and show that

$$\frac{1}{v_{ph}} - \frac{1}{v_{gr}} = \frac{\omega}{v_{ph}^2} \frac{dv_{ph}}{d\omega}$$

and that

$$v_{gr} = v_{ph} - \lambda \frac{dv_{ph}}{d\lambda}$$

Hint: Differentiate the modified expression for v_{ph} with respect to β, and $\beta = \frac{2\pi}{\lambda}$.

18.5 A transmission-line consists of two parallel strips of copper forming the go and return conductors, their width being six times the separation between them. The dielectric is air. From the Maxwell's equations applied to a TEM wave, show that the ratio of voltage to current in a progressive wave is 20π ohms.

Hint: Calculate L and C of the system, and hence the characteristic impedance.

18.6 Radar signals at 10,000 MHz are to be transmitted and recieved through a polystyrene window ($\varepsilon/\varepsilon_0 = 2.5$) let into the fuselage of an aircraft. Assuming that the waves are incident normally, how would you ensure that no reflections are produced by the window by choosing a particular thickness (= d) for the window?

Ans.: d = any integral multiple of 0.95 cm

18.7 By using Maxwell's equations, prove that a TEM wave cannot exist in a single conductor waveguide, such as rectangular or cylindrical waveguides.

19 Radiation and Reception of Electromagnetic Waves

19.1 INTRODUCTION

So far we have studied the propagation of electromagnetic waves in some detail. In Chapters 17 and 18, we studied their propagation in free space, in various dielectric and conducting media (extending to infinity), and then across the interfaces (i.e. planes of discontinuity) between different types of media, which produce the phenomena of reflection and refraction, and then along various guiding structures. But till now, we have not considered how and where such waves are produced and their effects at their destinations. So in the present chapter, we shall study the sources and the receivers of such waves.

The sources of any electromagnetic field are electric charges and currents. When these charges are time-varying, they may produce electromagnetic waves propagating away from the sources (and not returning to them). Such a process is known as 'the radiation of electromagnetic waves'. Theoretically, any time-varying sources of charges and currents radiate certain amount of energy. For example, at power frequency of 50 Hz, the radiation from the power transmission lines does exist in the rigorous sense, but in reality the radiated energy is so small that practically it cannot be detected, and hence justifiably considered as negligible.

In many practical problems, configurations of charges and currents radiating considerable energy to the surrounding space are required; and the science of designing and analyzing the 'radiating systems' has developed as a part of the theory of electromagnetism. Such a topic is highly specialized and complex. In our present introductory text, we intend to consider the basic, fundamental and simple types of radiating systems.

Systems designed for effectively radiating electromagnetic waves, are called antennae, of which there are many types. Once an antenna creates the em wave, the wave travels through the space to its destination where the message contained in the wave has to be extracted. The distance between the transmitting antenna and the receiving point is variable, and can be as much as hundreds of thousands of miles. For such large distances, special antennae which radiate very narrow beams are used. For receiving purposes, structures similar to transmitting antennae are used. These are called receiving antennae. There is a great diversity of such antennae, but their common characteristic is that they behave as generators when the incoming wave hits them

and a corresponding emf is induced in them. The analysis of receiving antennae is more complex than that of the transmitting antennae.

To design a radiating system, the information needed would include its wave pattern distribution, its polarization, and the distribution in space of the radiant energy, non-radiative losses, the currents and the charge distributions in the system, and its input impedance. These will depend on impressed frequency and voltage, the system dimensions, geometrical configuration, its materials and those of its surroundings. A complete solution of the problem must give fields which satisfy the Maxwell's equations and also the surface boundary conditions. There are methods of calculating the necessary required quantities with reasonable accuracy. By using the retarded potentials, the fields of charge and current distributions can be obtained. A rough estimate of the distant radiation (the far field) which is dependent on the major features of the source distributions is quite adequate. The ohmic losses can be calculated with sufficient accuracy by taking the real part of the complex Poynting vector.

19.2 THE ELECTRIC DIPOLE ANTENNA (THE HERTZIAN DIPOLE)

The electric dipole antenna or the Hertzian dipole (as it is more commonly known as) is the simplest of all the radiating systems, and is the basis for all types of antennae.

19.2.1 Electromagnetic Field of a Hertzian Dipole

The basic source of an electrostatic field is a point charge; and of magnetic field is a current loop. For an electromagnetic field, the simplest independent source is a *Hertzian Dipole* consisting of a pair of charges $\pm q$, separated by a distance l which is small compared with other distances considered; the charges vary with time, and are connected by a conductor in which is a current i which is

$$i = \frac{dq}{dt}$$

The electric field of the charges is associated, when the charges vary with time, with a displacement current which may be considered to complete the circuit for the element (i, l); alternatively, in terms of conduction currents only, we have a flow of charges which is associated with variation in the positive and negative charges at the termini (Figure 19.1).

Let us consider the field in free space surrounding such a dipole, in which (say) the charges and the current are given by

$$q = f(t) \qquad \text{and} \qquad i = f'(t) \tag{19.1}$$

In Section 13.5, we considered the magnetic vector potential applied to time-varying fields. Therein we have shown that the time-varying fields expressed by the Maxwell's equations can be expressed by the equations in terms of the magnetic vector potential $\mathbf{A}$ and a scalar electric potential V under the 'Lorentz gauge condition', such that the potential equations are completely separate. [Refer to Eqs. (13.52) to (13.54)]. These general equations reduce to the following form in free space (which we are considering at present), i.e.

$$\nabla \cdot \mathbf{A} + \left(\frac{1}{c^2}\right)\left(\frac{\partial V}{\partial t}\right) = 0 \tag{19.2a}$$

$$\nabla^2 \mathbf{A} - \left(\frac{1}{c^2}\right)\left(\frac{\partial^2 \mathbf{A}}{\partial t^2}\right) = 0 \tag{19.2b}$$

$$\nabla^2 V - \left(\frac{1}{c^2}\right)\left(\frac{\partial^2 V}{\partial t^2}\right) = 0 \tag{19.2c}$$

where $c = (\mu_0 \varepsilon_0)^{-1/2}$, as earlier in connection with the plane waves.

For a current element (i, l) under static conditions, we saw that **A** was a vector of magnitude $(\mu_0/4\pi)(il/r)$, parallel to the current element at every point [Ref.: Eq. (13.10) of Section 13.3.1]. It is not to be assumed that this will hold good under non-static conditions, but it will be assumed that the non-static solution still retains two features of the static, i.e.

(a) **A** is everywhere parallel to the element; and
(b) **A** depends only on the distance r from the element and on time t.

With these assumptions, **A** in Eqs. (19.2) has only one component which may be called A_z or just A.

Also,

$$\nabla \cdot \mathbf{A} = \frac{\partial A_z}{\partial z} = \frac{\partial A}{\partial z}$$

and

$$\left|\nabla^2 A\right| = \left(\frac{\partial^2 A}{\partial r^2}\right) + \left(\frac{2}{r}\right)\frac{\partial A}{\partial r} = \left(\frac{1}{r}\right)\left(\frac{\partial^2}{\partial r^2}\right)(rA)$$

These two equations may be proved by considering the flux out of an elementary volume in spherical coordinates.

Moreover with **A** being a function of r only, the maximum rate of change is $\left(\frac{\partial A}{\partial r}\right)$, and $\frac{\partial A}{\partial z} = \left(\frac{\partial A}{\partial r}\right)\cos\theta$.

The first two equations of the set (19.2), i.e. (b) and (a) therefore become in the reverse order as mentioned above:

$$\left(\frac{1}{r}\right)\left(\frac{\partial^2}{\partial r^2}\right)(rA) - \left(\frac{1}{c^2}\right)\left(\frac{\partial^2 A}{\partial t^2}\right) = 0 \tag{19.3}$$

and

$$\cos\theta\left(\frac{\partial A}{\partial r}\right) + \left(\frac{1}{c^2}\right)\left(\frac{\partial V}{\partial t}\right) = 0 \tag{19.4}$$

It may be verified that Eq. (19.3) is satisfied by

$$A = \left(\frac{1}{r}\right)F\left(t - \frac{r}{c}\right) \tag{19.5}$$

where F is any function.

$$\left[\text{This is because: } \left(\frac{1}{r}\right)\left(\frac{\partial^2}{\partial r^2}\right)(rA) = \left(\frac{1}{rc^2}\right)F'' = \left(\frac{1}{c^2}\right)\left(\frac{\partial^2 A}{\partial t^2}\right)\right]$$

We now compare Eq. (19.5) with the static equation

$$A = \left(\frac{\mu_0}{4\pi}\right)\left(\frac{il}{r}\right), \text{ which now becomes } = \frac{\mu_0 l f'(t)}{4\pi r}$$

when $i = f'(t)$ [Ref.: Eq. (19.1)].

Now c is a large number; thus for small values of r and moderate values of t, $F\left(t - \frac{r}{\mathrm{c}}\right)$ approximates to $F(t)$. If now we make

$$F(t) = \left(\frac{\mu_0 l}{4\pi}\right) f'(t)$$

then we see that Eq. (19.5) reduces to the static equation (or more accurately, the quasi-static). Hence the solution (19.5), for the more general case, is to be written as

$$A = \frac{\mu_0}{4\pi}\left(\frac{l}{r}\right) f'\left(t - \frac{r}{\mathrm{c}}\right) \tag{19.6}$$

Substitution of this value of A in Eq. (19.4) now gives

$$\frac{\partial V}{\partial t} = -\,\mathrm{c}^2 \cos\theta \left(\frac{\partial A}{\partial r}\right)$$

$$= \frac{\mu_0}{4\pi} l\mathrm{c}^2 \cos\theta \left[\left(\frac{1}{r^2}\right) f'\left(t - \frac{r}{\mathrm{c}}\right) + \left(\frac{1}{\mathrm{c}r}\right) f''\left(t - \frac{r}{\mathrm{c}}\right)\right] \tag{19.7}$$

whence

$$V = \frac{l\cos\theta}{4\pi\varepsilon_0}\left[\left(\frac{1}{r^2}\right) f\left(t - \frac{r}{\mathrm{c}}\right) + \left(\frac{1}{\mathrm{c}r}\right) f'\left(t - \frac{r}{\mathrm{c}}\right)\right] \tag{19.8}$$

Now $f(t - r/\mathrm{c})$ is the value of the charges q at an instant r/c earlier than the instant at which the value of V is being calculated; similarly $f'(t - r/\mathrm{c})$ is the value of i at an earlier instant. We write these as $[q]$, $[i]$, and call them the *retarded values* of q and i (refer to Section 13.6 on Retarded Potentials). Physically we see that this means that a signal travels from the point O to the point P with a finite velocity c, so that at the time t, an observer at P can only know about the values which q and i had at an earlier instant $(t - r/\mathrm{c})$.

Thus the potentials due to the oscillating dipole are

$$A = \left(\frac{\mu_0 l}{4\pi}\right)\left\{\frac{[i]}{r}\right\} \tag{19.9}$$

$$V = \left(\frac{l\cos\theta}{4\pi\varepsilon_0}\right)\left\{\frac{[q]}{r^2} + \frac{[i]}{\mathrm{c}r}\right\} \tag{19.10}$$

It should be noted that we did not assume the retarding effect to start with, but proved this to be an essential requirement for the time-varying nature of the radiating sources. If the 'retarding effect' is assumed, then these potentials can be obtained more directly (see Carter: *The Electromagnetic Field*, pp. 301–302). It must however be noted carefully that the field components **E** and **B** *cannot* be obtained by inserting the retarded values of q and i into the static expressions of **E** of the static dipole. It is essential to note that we started from the

potentials (and *not* the field expressions) which were shown to require the retarding effects to account for the time-varying nature of the sources.

Note: In our evaluation of the expressions for the two potentials, we have so far disregarded (or to be more precise, we have not been required to use it in our derivations so far) the third of the equations (19.2), i.e. Eq. (19.2c). If we use the Cartesian coordinate system, then:

$$\nabla^2 \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

The first equation (19.2a) is

$$\left(\frac{\partial \mathbf{A}}{\partial z}\right) + \left(\frac{1}{c^2}\right)\left(\frac{\partial V}{\partial t}\right) = 0 \tag{19.2a}$$

The second equation is

$$\left[\nabla^2 - \left(\frac{1}{c^2}\right)\left(\frac{\partial^2}{\partial t^2}\right)\right]\mathbf{A} = 0 \tag{19.2b}$$

Differentiating Eq. (19.2b) with respect to z, we get

$$\left[\nabla^2 - \left(\frac{1}{c^2}\right)\left(\frac{\partial^2}{\partial t^2}\right)\right]\left(\frac{\partial \mathbf{A}}{\partial z}\right) = 0 \tag{19.2b′}$$

Combining Eqs. (19.2a) and (19.2b′), we get

$$\left[\nabla^2 - \left(\frac{1}{c^2}\right)\left(\frac{\partial^2}{\partial t^2}\right)\right]\left(\frac{\partial V}{\partial t}\right) = 0$$

In other words, Eq. (19.7) is a solution of the equation, and V in Eq. (19.8) is of just the same form except for the different functions of $(t - r/c)$. Hence, we find that V, deduced as before, automatically satisfies Eq. (19.2c).

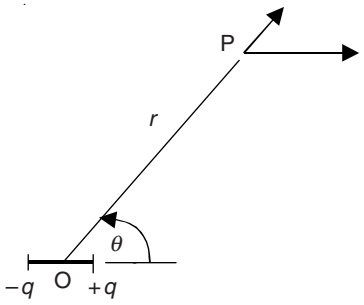


Figure 19.1 An oscillating dipole.

19.2.2 Field Components for the Hertzian Dipole

We use the spherical polar coordinate system with the origin at the centre of the dipole, and the angle θ measured from the axis of the dipole (Figure 19.2). In this coordinate system, a point P is defined by (r, θ, ϕ), and the element of length has the components—δr, $r\delta\theta$, $r \sin \theta \delta\phi$—and

$$\text{grad } V = \nabla V = \mathbf{i}_r\left(\frac{\partial V}{\partial r}\right) + \mathbf{i}_\theta\left(\frac{1}{r}\right)\left(\frac{\partial V}{\partial \theta}\right) + \mathbf{i}_\phi\left(\frac{1}{r \sin \theta}\right)\left(\frac{\partial V}{\partial \phi}\right)$$

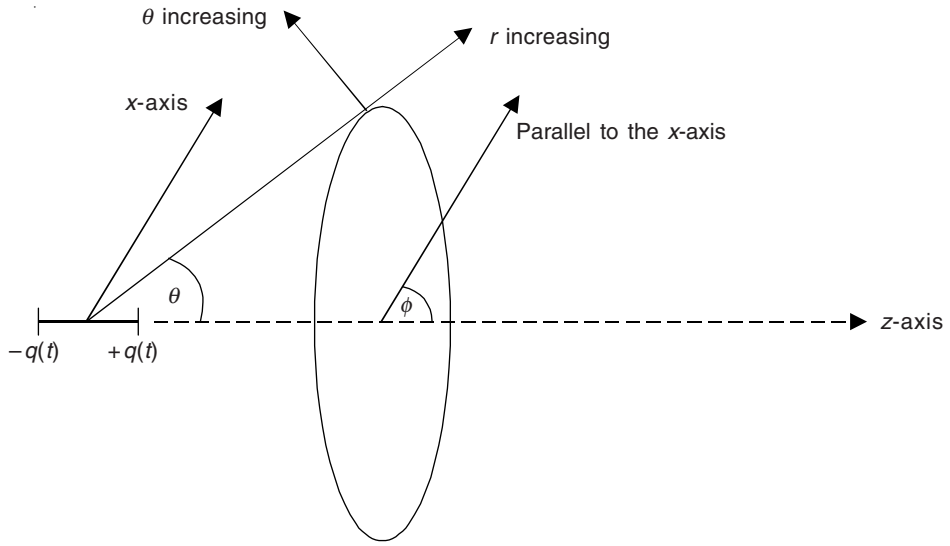


Figure 19.2 Oscillating dipole with spherical polar coordinate system.

In the present problem, V varies with r and θ, but not with ϕ [see Eq. (19.8)].

$\therefore$ (grad V) has r and θ components only.

Now,
$$\mathbf{E} = -\left(\frac{\partial \mathbf{A}}{\partial t}\right) - \nabla V$$

where $\mathbf{A}$ and V are given by Eqs. (19.6) and (19.8).

Also, $A_r = A\cos\theta$, and $A_\theta = A\sin\theta$.

$$\therefore \quad E_r = -\left(\frac{\partial A}{\partial t}\right)\cos\theta - \left(\frac{\partial V}{\partial r}\right)$$

$$= \left(\frac{\mu_0 l}{4\pi r}\right)\cos\theta f'' - \left(\frac{l\cos\theta}{4\pi\varepsilon_0}\right)\left[-\left(\frac{2}{r^3}\right)f - \left(\frac{2}{\mathrm{c}r^2}\right)f' - \left(\frac{1}{\mathrm{c}^2 r}\right)f''\right]$$

where $(t - r/\mathrm{c})$ is understood as the argument of each function; and if we make the substitution: $\mu_0 = 1/(\mathrm{c}^2\varepsilon_0)$, we find

$$E_r = \left(\frac{l\cos\theta}{4\pi\varepsilon_0}\right)\left[-\left(\frac{2}{r^3}\right)f\left(t - \frac{r}{\mathrm{c}}\right) - \left(\frac{2}{\mathrm{c}r^2}\right)f'\left(t - \frac{r}{\mathrm{c}}\right)\right] \tag{19.11}$$

Similarly,

$$E_\theta = -\left(\frac{\partial A}{\partial t}\right)\sin\theta - \left(\frac{1}{r}\right)\left(\frac{\partial V}{\partial \theta}\right)$$

$$= \left(\frac{\mu_0 l}{4\pi r}\right)\sin\theta f'' + \left(\frac{l\sin\theta}{4\pi\varepsilon_0 r}\right)\left[\left(\frac{1}{r^2}\right)f + \left(\frac{1}{\mathrm{c}r}\right)f'\right]$$

$$= \left(\frac{l\sin\theta}{4\pi\varepsilon_0}\right)\times\left[\left(\frac{1}{r^2}\right)f\left(t - \frac{r}{\mathrm{c}}\right) + \left(\frac{1}{\mathrm{c}r^2}\right)f'\left(t - \frac{r}{\mathrm{c}}\right) + \frac{1}{\mathrm{c}^2 r}f''\left(t - \frac{r}{\mathrm{c}}\right)\right] \tag{19.12}$$

The lines of force of the magnetic field form circles round the axis of the current (carried by the dipole). Thus B has a ϕ-component only. To evaluate B_ϕ, we consider the contour ABCDA as shown below in Figure 19.3:

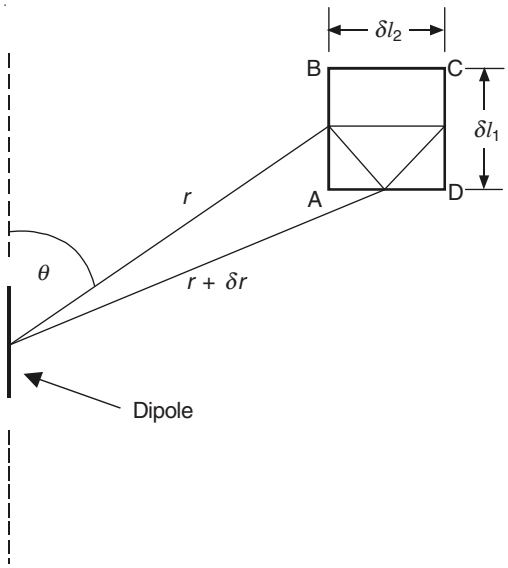


Figure 19.3 Contour for evaluating B_ϕ due to the dipole.

$$B_\phi = \frac{\text{line-integral of } \mathbf{A} \text{ round ABCDA}}{\text{area ABCD}}$$

$$= \frac{\delta l_1 \left[A(r) - A(r + \delta r) \right]}{\delta l_1 \delta l_2} = \left(\frac{1}{\delta l_2} \right) \left(-\frac{\partial A}{\partial r} \right) \delta r$$

Also, $\delta r = \delta l_2 \sin \theta$

$$\therefore \quad \mathbf{B}_\phi = -\left(\frac{\partial \mathbf{A}}{\partial r} \right) \sin \theta$$

Also, $B_\phi = (\nabla \times \mathbf{A})_\phi = \dfrac{\partial A_\theta}{\partial r} + \dfrac{A_\theta}{r} - \left(\dfrac{1}{r} \right) \left(\dfrac{\partial A_r}{\partial \theta} \right)$

with $A_r = A \cos \theta$, $A_\theta = -A \sin \theta$, where A is a function of r only (and of t).

$$\therefore \qquad B_\phi = \sin \theta \left(-\frac{\partial A}{\partial r} - \frac{A}{r} + \frac{A}{r} \right)$$

$$= -\left(\frac{\partial A}{\partial r} \right) \sin \theta$$

Thus,

$$B_\phi = \left(\frac{\mu_0 l}{4\pi}\right)\sin\theta\left[\left(\frac{1}{r^2}\right)f'\left(t-\frac{r}{\mathrm{c}}\right)+\left(\frac{1}{\mathrm{c}r}\right)f''\left(t-\frac{r}{\mathrm{c}}\right)\right] \tag{19.13}$$

If (r/c) is negligible, the retardation is of no effect, and Eqs. (19.11), (19.12), and (19.13) can be written as

$$E_r = \frac{ql\cos\theta}{2\pi\varepsilon_0 r^3} \tag{19.14a}$$

$$E_\theta = \frac{ql\sin\theta}{4\pi\varepsilon_0 r^3} \tag{19.14b}$$

$$B_\phi = \frac{\mu_0 il\sin\theta}{4\pi r^2} \tag{19.14c}$$

These are the equations for a static dipole, and steady current element respectively.

At a great distance, however, the terms containing ($1/r$) predominate, and we get (the far field):

$$E_r = 0 \tag{19.15a}$$

$$E_\theta = \left(\frac{l\sin\theta}{4\pi\varepsilon_0 \mathrm{c}^2 r}\right)f''\left(t-\frac{r}{\mathrm{c}}\right)$$

$$= \left(\frac{\mu_0 l\sin\theta}{4\pi r}\right)f''\left(t-\frac{r}{\mathrm{c}}\right) \tag{19.15b}$$

$$B_\phi = \left(\frac{\mu_0 l\sin\theta}{4\pi r\mathrm{c}}\right)f''\left(t-\frac{r}{\mathrm{c}}\right) \tag{19.15c}$$

Thus the field at this distance consists of electric and magnetic components, perpendicular to each other, and both perpendicular to the radius, and related by the simple relationship:

$$E_\theta = \mathrm{c}B_\phi \tag{19.16}$$

This pattern travels outwards with a velocity c, diminishing in proportion to ($1/r$); it forms, in fact, 'a travelling electromagnetic wave'. It is this waveform, which made the radio-telegraphy and telephony possible. The Poynting vector shows that it is associated with an outward flow of energy into space.

The Hertzian dipole is not in itself an efficient form of transmitting aerial (or antenna), but its field is the 'brick' from which the fields of practical aerials are built up.

19.2.3 Radiated Power of the Hertzian Dipole Calculated by the Poynting Vector, and the Radiation Resistance

We now apply the Poynting vector to find the energy transmitted from an oscillating dipole. As in the previous sections, the dipole is made up of the charges $\pm q$, where $q = f(t)$, and the charges are separated by a distance l.

At a *great* distance r from the dipole, the electric field was found to be in the θ-direction (i.e. a line of longitude on a sphere concentric with the dipole), and given by the expression:

$$E_\theta = \left(\frac{\mu_0 l}{4\pi}\right)\left(\frac{\sin\theta}{r}\right) f''\left(t - \frac{r}{c}\right) \tag{19.15b}$$

and the magnetic field lies along a parallel of a latitude, i.e.

$$B_\phi = \left(\frac{\mu_0 l}{4\pi c}\right)\left(\frac{\sin\theta}{r}\right) f''\left(t - \frac{r}{c}\right) \tag{19.15c}$$

$\therefore$ The Poynting vector is radially outwards,

$$\mathbf{S} = \mathbf{E} \times \mathbf{H}, \quad \text{where } H_\phi = \frac{B_\phi}{\mu}$$

And hence the magnitude of the Poynting vector is

$$S = \left(\frac{\mu_0 l^2}{16\pi^2 c}\right)\left(\frac{\sin\theta}{r}\right)^2 \left[f''\left(t - \frac{r}{c}\right)\right]^2 \tag{19.17}$$

i.e. a travelling wave of energy in which the direction of flow is never reversed (like in a coaxial cable).

To obtain the total outflow of power from the dipole, we have to integrate over the whole surface of the sphere, concentric with the dipole, as shown in Figure 19.4.

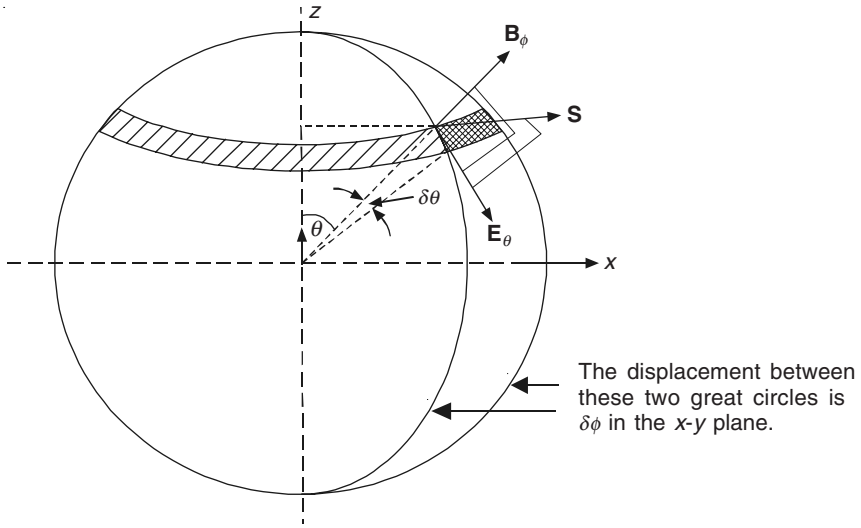


Figure 19.4 Surface of integration around the dipole. Note that the three vectors $\mathbf{E}_\theta$, $\mathbf{B}_\phi$ and $\mathbf{S}$ are mutually orthogonal.

Area of the cross-hatched curvilinear rectangle between the angles θ and $\theta + \delta\theta$, and $\delta\phi$ is $(r\,\delta\theta)(r \sin\theta\,\delta\phi)$.

$\therefore$ Area of the circular strip (shown dashed) between the angles θ and $\theta + \delta\theta$, is $(2\pi r \sin\theta)(r\,\delta\theta)$.

For the complete area covering the whole surface of the enclosing sphere, θ varies from $\theta = 0$ to $\theta = \pi$.

Total power, $$W = \int_{\theta=0}^{\theta=\pi} S(2\pi r \sin\theta)(r\,\delta\theta)$$

$$= \left(\frac{\mu_0 l^2}{16\pi^2 \mathrm{c}}\right)\left[f''\left(t-\frac{r}{\mathrm{c}}\right)\right]^2 \int_0^{\pi}\left(\frac{\sin\theta}{r}\right)^2 2\pi r^2 \sin\theta\, d\theta$$

$$= \left(\frac{\mu_0 l^2}{8\pi \mathrm{c}}\right)\left[f''\left(t-\frac{r}{\mathrm{c}}\right)\right]^2 \int_0^{\pi} \sin^3\theta\, d\theta$$

In this integral, substitute $\cos\theta = x$; $\therefore -\sin\theta\, d\theta = dx$, and the limits are: $\theta = 0 \Rightarrow x = 1$; and $\theta = \pi \Rightarrow x = -1$.

$$\int_0^{\pi} \sin^3\theta\, d\theta = \int_{-1}^{+1}(1-x^2)\,dx = \left(x - \frac{x^3}{3}\right)_{-1}^{+1} = \left(2 - \frac{2}{3}\right) = \frac{4}{3}$$

$\therefore$ $$W = \left(\frac{\mu_0 l^2}{6\pi c}\right)\left[f''\left(t-\frac{r}{\mathrm{c}}\right)\right]^2$$

Usually $f(t)$ is a function varying sinusoidally with time. Hence, we can write:

$$q = f(t) = Q_{\mathrm{m}} \sin\omega t, \text{ where } Q_{\mathrm{m}} = \text{amplitude of the charge}$$

$\therefore$ $$f''\left(t-\frac{r}{\mathrm{c}}\right) = -\omega^2 Q_{\mathrm{m}} \sin\left[\omega\left(t-\frac{r}{\mathrm{c}}\right)\right]$$

$\therefore$ The total power, $$W = \left(\frac{\mu_0 l^2 \omega^4 Q_{\mathrm{m}}^2}{6\pi \mathrm{c}}\right)\sin^2\left[\omega\left(t-\frac{r}{\mathrm{c}}\right)\right] \tag{19.18}$$

$\therefore$ W varies sinusoidally between 0 and 1.

To find the mean output of power, we have to consider the time-varying term and integrate with respect to t over one time-period. The mean value of this integrated term is (1/2).

$\therefore$ The mean power, $$\bar{W} = \left(\frac{\mu_0 l^2 \omega^4}{6\pi \mathrm{c}}\right)\left(\frac{Q_{\mathrm{m}}^2}{2}\right)$$

$$= \frac{\mu_0 l^2 \omega^4 Q^2}{6\pi \mathrm{c}} \tag{19.19}$$

where $\left(\frac{Q_{\mathrm{m}}}{\sqrt{2}}\right) = Q \leftarrow$ rms value of the time-varying charge.

In the transmitting aerials, it is more common to use the current than the charge. Hence:

The rms (= root mean square) current, $I = \omega Q$

$\therefore$ The mean power, $$\bar{W} = \frac{\mu_0 l^2 \omega^2 I^2}{6\pi \mathrm{c}} \tag{19.20}$$

We substitute for the angular frequency ω,

$$\omega = \frac{2\pi c}{\lambda}, \quad \text{where } \lambda \text{ is the wavelength}$$

$\therefore$ The mean power, $$\bar{W} = \left(\frac{\mu_0 l^2}{6\pi c}\right)\left(\frac{4\pi^2 c^2}{\lambda^2}\right) I^2$$

$$= \left(\frac{2}{3}\right)\pi Z_0 \left(\frac{l^2}{\lambda^2}\right) I^2 = R_e I^2 \tag{19.21}$$

where $\mu_0 c = \sqrt{\mu_0/\varepsilon_0} = Z_0$ (= wave impedance of free space).

R_e (= R_{rad}) is called the *radiation resistance* of the aerial at the wavelength in question.

$$\therefore \quad R_e = \left(\frac{2}{3}\right)\pi Z_0 \left(\frac{l^2}{\lambda^2}\right) \tag{19.22}$$

But, $Z_0 = 377\ \Omega$ $\quad \therefore \quad R_e = 790\,\frac{l^2}{\lambda^2}$.

Thus the Hertzian dipole is rather inefficient because of the $(l/\lambda)^2$ term as $l << \lambda$.

The above expression explains how it is possible for a transmitting aerial to send out power into the surrounding space, and how the process of transmission reacts back upon the source by bringing about an apparent increase in the resistance.

So the Hertz dipole can be considered as a resistance when seen from the input terminals of the antenna [Figure 19.4(a)].

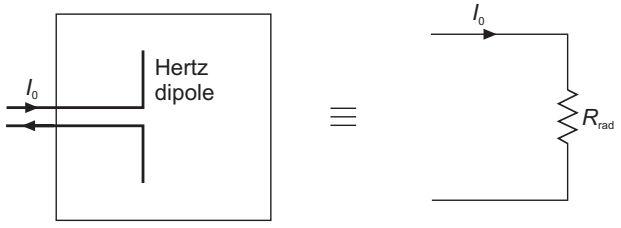


Figure 19.4(a) Equivalent resistance of the Hertz dipole.

This is quite low for a Hertzian dipole as mentioned above, e.g., for a 0.1λ long dipole, $R_{rad} = 8\ \Omega$ only

Note that E_θ and B_ϕ are in time-phase; and I is not necessarily a constant, but a function of z.

We shall see later that $l = \lambda/2$ is a more efficient choice of an antenna.

19.2.4 Distributions of Potentials, Field Vectors, and the Poynting Vector of the Hertzian Dipole

We have obtained the expressions for the vector and the scalar potentials of the Hertzian dipole (i.e. the oscillating electric dipole) in Section 19.2.1, as:

$$\mathbf{A} = \mathbf{i}_z \left(\frac{\mu_0}{4\pi}\right)\left(\frac{I_0 l}{r}\right)\exp\left[j\omega\left(t - \frac{r}{\mathrm{c}}\right)\right] \tag{19.9′}$$

$$V = \left(\frac{l\cos\theta}{4\pi\varepsilon_0}\right)\times\left[\left(\frac{Q_0}{r^2}\right)\exp\left\{j\omega\left(t - \frac{r}{\mathrm{c}}\right)\right\} + \left(\frac{I}{\mathrm{c}r}\right)\exp\left\{j\omega\left(t - \frac{r}{\mathrm{c}}\right)\right\}\right] \tag{19.10′}$$

The plots of **A** and V are shown in Figure 19.5. It is seen that the scalar potential V is a maximum at the poles and becomes zero at the equator, where the potentials due to the individual charges $\pm Q$ of the dipole cancel each other. The charge of $+Q$ predominates in the northern hemisphere, and that of $-Q$ in the southern hemisphere. The magnitude of **A** is represented by the sphere marked A (Figure 19.5), and is independent both in direction and magnitude, of the coordinates θ and ϕ.

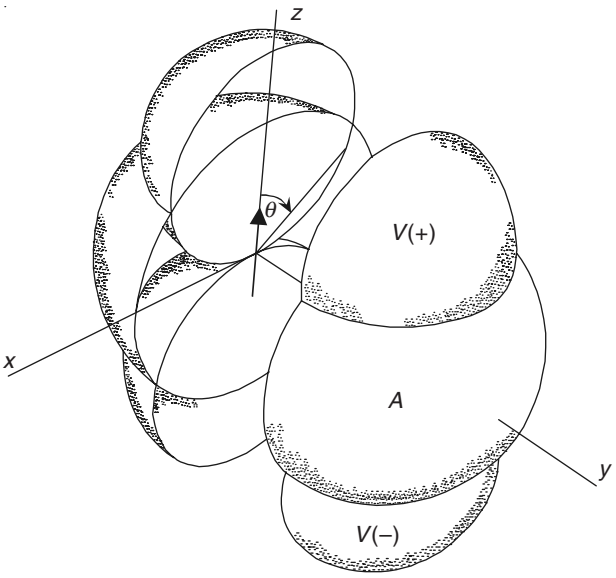


Figure 19.5 The scalar potential V and the magnitude of the magnetic vector potential, i.e. $|\mathbf{A}|$, shown as functions of θ and ϕ about the dipole oriented as shown.

The electric lines of force can be obtained on the basis of the consideration that when (r/c) is negligible, the retardation effects can be neglected, and then the **E** vector [i.e. Eqs. (19.11) and (19.12)] collapses to the expressions of the static dipole. Then,

$$\frac{E_r}{dr} = \frac{E_\theta}{r\,d\theta} \tag{19.14′}$$

since an element of a line of force having the components dr and $rd\theta$ is parallel to the local **E** having the components E_r and E_θ. To simplify the calculations, an auxiliary vector **C** can be used such that:

$$\mathbf{C} = \mathbf{i}_\phi \left(\frac{Q_0 l}{4\pi\varepsilon_0}\right)\left(\frac{l}{r} + \frac{j\omega}{\mathrm{c}}\right)\exp\left[j\left(\omega t - \frac{r}{\mathrm{c}}\right)\right]\sin\theta \tag{19.14a′}$$

whence $\mathbf{E} = \nabla \times \mathbf{C}$, so that

$$E_r = \left(\frac{1}{r \sin \theta}\right)\frac{\partial}{\partial \theta}(C \sin \theta) \tag{19.14b′}$$

and

$$E_\theta = \left(-\frac{1}{r}\right)\frac{\partial}{\partial r}(rC) \tag{19.14c′}$$

The differential equation for the lines of force is therefore:

$$\left(\frac{1}{\sin \theta}\right)\frac{\partial}{\partial \theta}(C \sin \theta) = \left(-\frac{1}{r}\right)\frac{\partial}{\partial r}(rC)$$

or

$$\left[\frac{\partial}{\partial \theta}(Cr \sin \theta)\right]d\theta + \left[\frac{\partial}{\partial r}(Cr \sin \theta)\right]dr = 0$$

$$\therefore \quad Cr \sin \theta = \text{constant} \tag{19.14d′}$$

Figure 19.6 shows four families of lines of force of **E** at four instants of time. The magnetic lines of force are much simpler. They are circles, perpendicular to, and centred on the axis of the dipole.

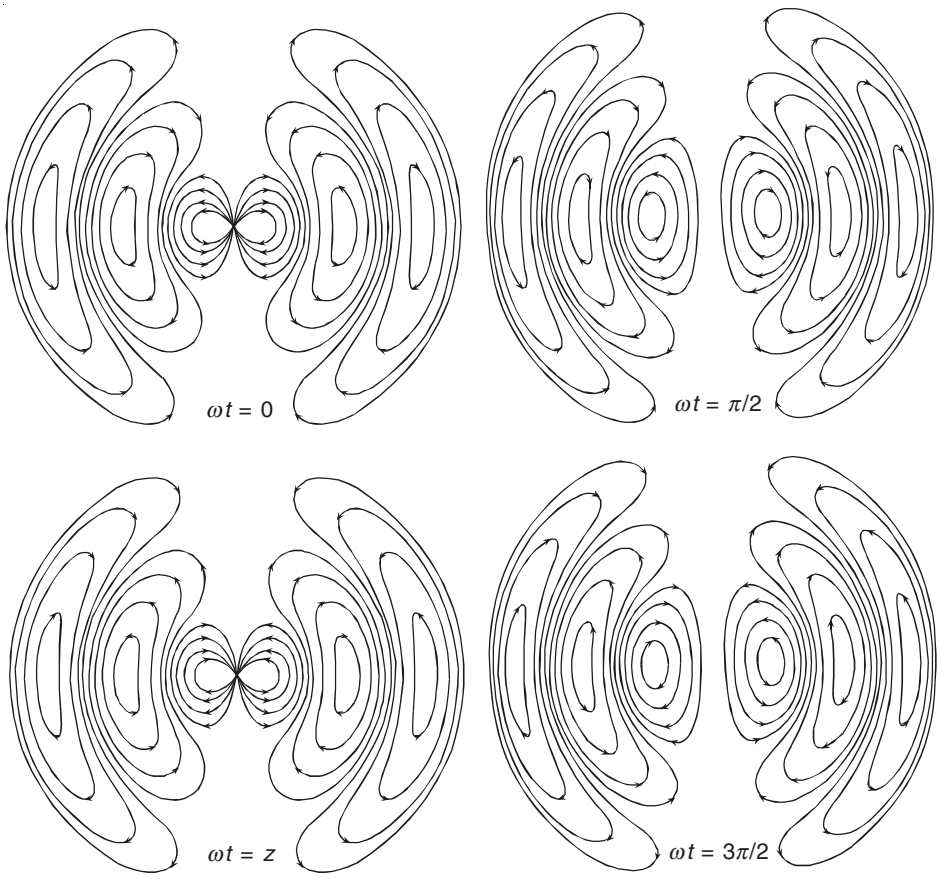


Figure 19.6 Lines of **E** of a Hertzian dipole for $\omega t = 0, \pi/2, \pi, 3\pi/2$. The dipole is situated at the centre and oriented vertically.

Figure 19.7 shows the polar diagrams (or the angular distributions) of **E** or **H**, and $\mathbf{S}_{av}$ at distances much greater than the dipole wavelength λ. In fact, Eqs. (19.14) and (19.17) show that **E** and **S** vary as sin θ and $\sin^2\theta$ respectively; and hence these are effectively plots of these two functions. The Poynting vector varies as $(1/r^2)$, and **E** and **B** vary as $(1/r)$. Since the energy flow varies as $\sin^2\theta$, it is zero along the axis of the dipole, and is maximum in the equatorial plane. So *an electric dipole does not radiate energy along its axis*. Also, the energy flow is radial everywhere.

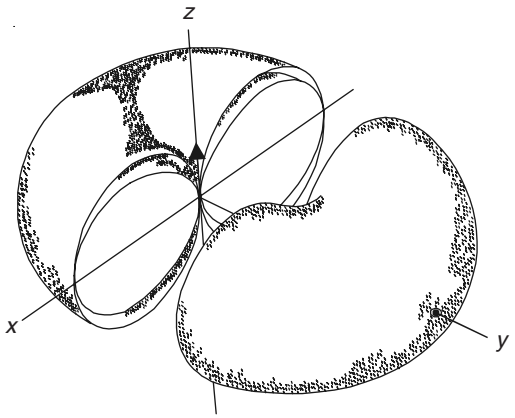


Figure 19.7 Plots of angular distributions of **E** (or **H**) and $\mathbf{S}_{av}$ of the Hertzian dipole.

19.2.5 Thin Linear Antenna

The Hertz dipole which is a very small current element has a very small radiation resistance, and hence is not an efficient radiator. So a logical extrapolation would be a dipole of reasonable physical length, such that the antenna can then be used for both transmission and reception of the electromagnetic radiation.

A linear dipole is then a piece of wire of length $2L$ (say) excited by a current or voltage source as shown in Figure 19.7(a).

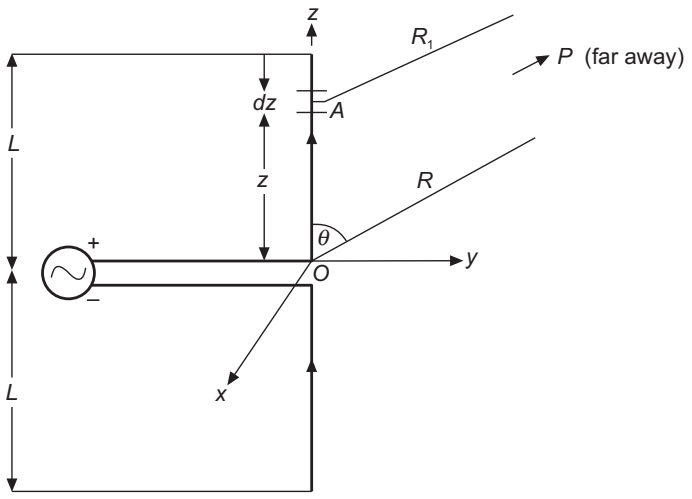


Figure 19.7(a) A linear dipole antenna of length $2L$.

The gap at the centre of the dipole is assumed to be small, and the current spreads over its length only to achieve a steady-state value (or distribution). Finding the current distribution along the antenna is not an easy problem. Once the current distribution is known, the radiation characteristic of the antenna can be predicted. It can be argued that where the wire ends the current will be zero and this antenna can be visualized as a flared-up version of a twin-wire transmission line as shown in Figure 19.7(b).

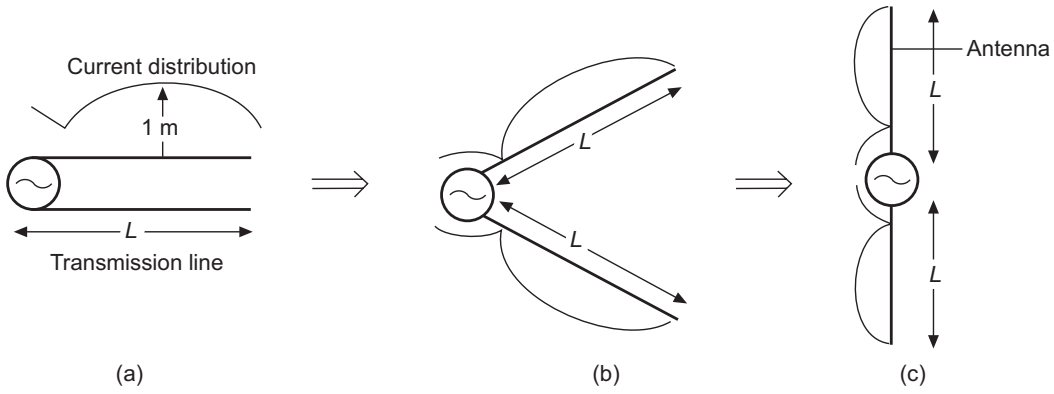


Figure 19.7(b) A dipole antenna as a flared version of a transmission line.

We consider an o.c. section of a transmission line (twin-wire types of length L). The voltage reflection coefficient at the o.c. end is +1, and hence there would be a standing wave along the line. The current at the o.c. end would be zero and its magnitude varies sinusoidally along the length.

The standing wave current distribution on the line

$$\left.\begin{aligned} I(z) &= I_m \sin[\beta(L - |z|)] \\ &= I_m \sin[\beta(L - z)], \quad \text{for } z > 0 \\ &\quad I_m \sin[\beta(L + z)], \quad \text{for } z < 0 \end{aligned}\right\} \tag{19.22a}$$

where

I_m = maximum current amplitude on the line

$\beta = \dfrac{2\pi}{\lambda}$, i.e. the propagation constant in free space

z = distance measured on the line from the generator, +ve on the upper conductor and –ve on the lower conductor.

As the line becomes flared, it becomes a non-uniform transmission line. When the line becomes an antenna, we assume the current distribution to remain unchanged, which rigorously speaking is a questionable assumption. For the analysis, we simplify the current distribution to be sinusoidal [$I \sin(\omega t - \beta r)$]. To evaluate the field due to this linear dipole, the dipole is sub-divided into infinitesimally small current elements, and obtain the field due to each element and then apply the principle of superposition to obtain the field due to the whole dipole. So considering an element dz at a height z with the coordinate system as shown in Figure 19.17(a), the E_θ field can be expressed as [referring to Eq. (19.15b) for the far field components]

$$dE_\theta = j\frac{\beta^2 \sin\theta}{4\pi\omega\varepsilon}\frac{I(z)\,dz\,e^{-j\beta R_1}}{R_1} \tag{19.22b}$$

and the associated magnetic field

$$dH_\theta = \frac{dE_\theta}{\eta} \tag{19.22c}$$

Since the point under consideration (i.e. P) is far away from the antenna, lines AP and OP are effectively parallel, and so

$$R_1 = R - z\cos\theta \tag{19.22d}$$

In the expression for dE_θ, R_1 in the denominator can be replaced by R but not in the exponent because it is considered relative to the wavelength λ. Hence

$$dE_\theta = j\frac{\beta^2 \sin\theta}{4\pi\omega\varepsilon R} I(z)\, e^{j(-\beta R + \beta z\cos\theta)}dz \tag{19.22e}$$

The total field due to the complete dipole will be

$$E_\theta = j\frac{\beta^2 \sin\theta\, e^{-j\beta R}}{4\pi\omega\varepsilon R}\int_{-L}^{L} I(z)\, e^{j\beta z\cos\theta}d\theta$$

$$= j60\, I_m \frac{e^{-j\beta R}}{R} F(\theta) \tag{19.22f}$$

where $I(z)$ is as stated earlier, and

$$\frac{\beta^2}{4\pi\omega\varepsilon} = \frac{\omega^2\mu\varepsilon}{4\pi\omega\varepsilon} = \frac{\omega\mu}{4\pi} = \frac{\beta\eta_0}{4\pi} = 30\beta \tag{19.22g}$$

and

$$F(\theta) = \frac{\cos(\beta L\cos\theta) - \cos(\beta L)}{\sin\theta} \tag{19.22h}$$

$F(\theta)$ describes the relative variation of the electric field as a function of θ and so it is the E-plane radiation pattern. In the H-plane, E_θ, not being a function of ϕ, is constant. So the H-plane radiation pattern of this dipole is same as that of the Hertzian dipole. But as seen the E-plane pattern is different from that of the Hertzian dipole.

19.3 THE HALF-WAVE ANTENNA

The oscillating dipole discussed so far, is a useful tool for theoretical work, but it is not a practical antenna. The half-wave antenna, shown in Figure 19.8, is simply a straight conductor whose length is half a free-space wavelength. It is fed at the centre (i.e. a current ($I_0 \cos \omega t$) is established at the centre of the dipole) by means of a suitable electronic circuit, and a standing wave is formed along the conductor.

Note: It should be noted that, in many cases, the half-wave antenna is a quarter wavelength mast set vertically on the ground which then behaves like a mirror. The quarter-wave mast and

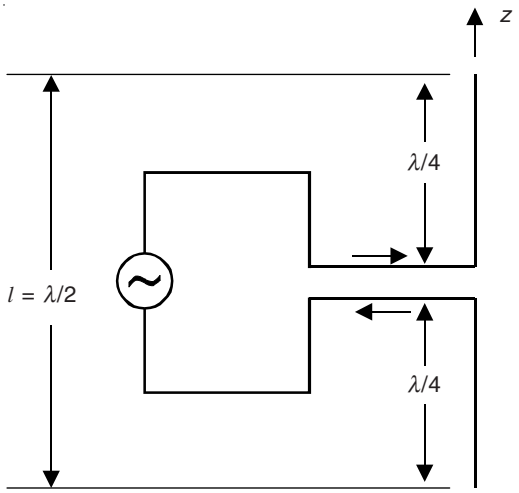


Figure 19.8 Centre-fed half-wave antenna.

its image in the ground then together form a half-wave antenna. Radio broadcast antennae are often of this type. The ground in the neighbourhood of the antenna can be covered with a conducting screen in order to achieve good conductivity.

The standing wave of the current on the conductor is

$$I = I_0 \exp(j\omega t) \cos(\beta l) \tag{19.23a}$$

Each element ($I\,dl$) of the antenna then radiates an electromagnetic wave similar to that of an electric dipole, and the field at any point in space is the sum of all these fields.

Since the (magnetic) vector potential at any point will be parallel to the direction of the source current in the antenna, the vector potential will have only the z-component (parallel to l).

The vector potential at the point P(r, ϕ, θ) due to the antenna element ($I\,dl$), at a distance l from the centre-point of the antenna (which is also the origin of our coordinate system) is given by

$$dA_z = \frac{\mu_0 I \exp(-j\beta R)\, dl}{4\pi R} \tag{19.23b}$$

where R is the distance between the point P and the antenna element ($I\,dl$). Since the distance R to the point of observation is $>> \lambda$ (we are interested in the far field only), we shall be able to make some simplifications in our analysis as we proceed.

The total vector potential $\mathbf{A}_z$ at P, due to all the current elements in the antenna is

$$A_z = \left(\frac{\mu_0}{4\pi}\right) \int_{-\lambda/4}^{0} I_0 \cos(\beta l) \exp(-j\beta R) \frac{dl}{R} + \left(\frac{\mu_0}{4\pi}\right) \int_{0}^{\lambda/4} I_0 \cos(\beta l) \exp(-j\beta R) \frac{dl}{R} \tag{19.23c}$$

In the denominators, we can approximate $R \approx r$, but for the R in the phase factor of the numerator, the difference between R and r is important. Also, for the large values of R, the two lines R (from the element dl to P) and r (from the centre (the origin) to P) can be considered as parallel lines, and we can write

$$R = r - l \cos\theta \tag{19.24}$$

$\therefore$ The above expression for $\mathbf{A}$ simplifies to

$$A_z = \left(\frac{\mu_0 I_0}{4\pi r}\right)\exp(-j\beta r) \times \int_0^{\lambda/4} \cos(\beta l)\left[\exp(j\beta l\cos\theta) + \exp(-j\beta l\cos\theta)\right] dl$$

$$\therefore \quad A_z = \left(\frac{\mu_0 I_0}{4\pi r}\right)\exp(-j\beta r) \times \int_0^{\lambda/4} \left[\cos\{\beta z(1+\cos\theta)\} + \cos\{\beta z(1-\cos\theta)\}\right] dl$$

$$= \left[\frac{\mu_0 I_0 \exp(-j\beta r)}{2\pi\beta r}\right]\left[\frac{\cos\left\{\left(\frac{\pi}{2}\right)\cos\theta\right\}}{\sin^2\theta}\right] \tag{19.25}$$

Since the current is entirely in the z-direction,

$$H_\phi = \left(\frac{1}{\mu_0}\right)\left[-\left(\frac{\partial A_z}{\partial r}\right)\sin\theta\right]$$

$$= \left[\frac{jI_0 \exp(-j\beta r)}{2\pi r}\right]\left[\frac{\cos\left\{\left(\frac{\pi}{2}\right)\cos\theta\right\}}{\sin\theta}\right] \tag{19.26}$$

retaining only $(1/r)$ terms for the larger distances considered.

The electric field strength,

$$E_\theta = Z_0 H_\phi = \left[\frac{j60 I_0 \exp(-j\beta r)}{r}\right]\left[\frac{\cos\left\{\left(\frac{\pi}{2}\right)\cos\theta\right\}}{\sin\theta}\right] \tag{19.27}$$

$\therefore$ The magnitude of the **E** field for the radiation field of a half-wave dipole (or a quarter-wave monopole) is

$$|E_\theta| = \left(\frac{60 I_0}{r}\right)\left[\frac{\cos\left\{\left(\frac{\pi}{2}\right)\cos\theta\right\}}{\sin\theta}\right] \quad \frac{\text{V}}{\text{m}} \tag{19.28}$$

For $\theta = 0$ and $\theta = \pi$, the expressions for E_θ and H_ϕ become indeterminate, because the trigonometric term becomes (0/0). To evaluate it (using l'Hospital's rule), we get

$$\lim_{\theta\to 0 \text{ or } \pi}\left[\frac{\cos\left\{\left(\frac{\pi}{2}\right)\cos\theta\right\}}{\sin\theta}\right] = \left.\frac{\frac{d}{d\theta}\left[\cos\left\{\left(\frac{\pi}{2}\right)\cos\theta\right\}\right]}{\frac{d}{d\theta}(\sin\theta)}\right|_{\theta\to 0 \text{ or } \pi}$$

$$= \left.\frac{\sin\left\{\left(\frac{\pi}{2}\right)\cos\theta\right\}\left(\frac{\pi}{2}\right)\sin\theta}{\cos\theta}\right|_{\theta\to 0 \text{ or } \pi} \tag{19.29}$$

$\therefore$ E is zero along the axis of the antenna.

E_θ and H_ϕ are in time-phase, and hence the maximum value (in time) of the Poynting vector is the product of the peak values of E_θ and H_ϕ.

$\therefore$ The average value in time of the Poynting vector (= 1/2 the peak value),

$$\mathbf{S}_{av} = \left(\frac{Z_0 I_0^2}{8\pi^2 r^2}\right)\left[\frac{\cos^2\left\{\left(\frac{\pi}{2}\right)\cos\theta\right\}}{\sin^2\theta}\right] \tag{19.30}$$

It points radially outwards, and varies as $(1/r^2)$, and hence ensures conservation of energy.

The total radiated power is obtained by integrating $\mathbf{S}_{av}$ over a sphere of radius r:

$$W = 9.55\,(I_{rms})^2 \int_0^{2\pi}\int_0^{\pi}\left[\frac{\cos^2\left\{\left(\frac{\pi}{2}\right)\cos\theta\right\}}{r^2\sin^2\theta}\right] r^2 \sin\theta\, d\theta\, d\phi$$

$$= 60\,(I_{rms})^2 \int_0^{\pi}\left[\frac{\cos^2\left\{\left(\frac{\pi}{2}\right)\cos\theta\right\}}{\sin\theta}\right] d\theta \tag{19.31}$$

by replacing I_0 by its rms value I_{rms}.

In most of the radiation problems, the main difficulty lies in the evaluation of some integral. The above integral can be evaluated in a number of ways. One possible way is to use the substitution,

$$\left(\frac{\pi}{2}\right)\cos\theta = \frac{\alpha}{2} - \frac{\pi}{2} \tag{19.32}$$

$\therefore$
$$I = \int_0^{\pi}\left[\frac{\cos^2\left\{\left(\frac{\pi}{2}\right)\cos\theta\right\}}{\sin\theta}\right] d\theta = \int_0^{2\pi}\left[\frac{1-\cos\alpha}{\alpha(4\pi-2\alpha)}\right] d\alpha \tag{19.33}$$

Breaking up the fraction:

$$\frac{1}{\alpha(4\pi-2\alpha)} = \left(\frac{1}{4\pi}\right)\left(\frac{1}{\alpha} + \frac{1}{2\pi-\alpha}\right) \tag{19.34}$$

$\therefore$
$$I = \left(\frac{\pi}{4\pi}\right)\left[\int_0^{2\pi}\left(\frac{1-\cos\alpha}{\alpha}\right) d\alpha + \int_0^{2\pi}\left(\frac{1-\cos\alpha}{2\pi-\alpha}\right) d\alpha\right]$$

$$= \left(\frac{1}{2}\right)\int_0^{2\pi}\left(\frac{1-\cos\alpha}{\alpha}\right) d\alpha = \left(\frac{1}{2}\right)\times 2.4377 \tag{19.35}$$

by Simpson's rule (numerically).

$$\therefore \quad W = 73.1(I_{rms})^2 \text{ watts} \tag{19.36}$$

$$R_{rad} = 73.1\ \Omega \tag{19.37}$$

Thus the radiation resistance of the half-wave antenna is 73.1 Ω, on the basis of the assumption that the current distribution on the antenna is sinusoidal, which however is not quite correct. This is because, the standing wave pattern assumed on the antenna can be truly sinusoidal, only if there is zero energy loss, and hence no radiated wave. But a more rigorous calculation shows nearly the same result as the present approximate idealized calculation. So the present simplified assumption is a justifiable one.

The radiated power by this antenna is also greater in the equatorial plane, compared with the dipole. If we plot the angular distributions of **E**, **H**, and the Poynting vector for the half-wave antenna, then the comparable functions are $[\cos\{(\pi/2)\cos\theta)\}/\sin\theta]$ for E and $[\cos^2\{(\pi/2)\cos\theta)\}/\sin^2\theta]$ for $\mathbf{S}_{av}$, as against $\sin\theta$ for E and $\sin^2\theta$ for $\mathbf{S}_{av}$ of the Hertzian dipole. The trigonometric functions for the half-wave antenna produce flatter polar diagrams compared with the torii produced by the $\sin\theta$ and $\sin^2\theta$ for the Hertzian dipole (refer to Figure 19.7). The polar diagrams for the half-wave antenna are plotted below in Figure 19.9.

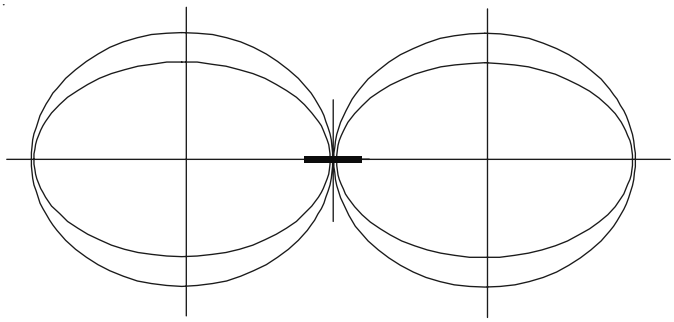


Figure 19.9 Angular plots of $[\cos\{(\pi/2)\cos\theta\}/\sin\theta]$ and $[\cos^2\{(\pi/2)\cos\theta\}/\cos^2\theta]$ showing respectively the angular distributions of **E** and $\mathbf{S}_{av}$ for a half-wave antenna at $r >> \lambda$. The plots are similar to those of the dipole (Figure 19.7) except that these are flatter compared to those of the dipole, indicating a larger fraction of its power in the equatorial plane.

19.3.1 Quarter-wave Monopole Antenna

Since at low frequencies, the wavelength becomes too large, the length of the dipole antenna also becomes large. So at medium frequencies (wavelength approximately hundreds of metres), the dipole antenna is not a viable choice. So a quarter-wave monopole antenna consisting of one-half of a half-wave dipole, normal to a conducting ground plane is used. The base plane is assumed to be infinite and perfectly conducting. A co-axial cable connected to its base, feeds it.

A monopole antenna of length L is shown here, and a voltage has been applied between the bottom of the antenna and the ground. By using the image theory, the infinite, perfectly conducting ground plane is replaced by the image of the monopole. The field produced in the region above the ground plane, by the $\lambda/4$ monopole and its image is same as the field produced by a $\lambda/2$ dipole. For calculating the radiated power, the limit of integration in Eq. (19.31) should be replaced by $\theta = 0$ to $\theta = \pi/2$, as the monopole radiates only through the upper hemispherical surface. Thus the monopole radiates only half the power radiated by the dipole with the same current. Hence

$$P_{rad} = 18.28 I_0^2 \quad \text{and} \quad R_{rad} = \frac{2P_{rad}}{I_0^2} \quad \text{or} \quad R_{rad} = 36.5\ \Omega$$

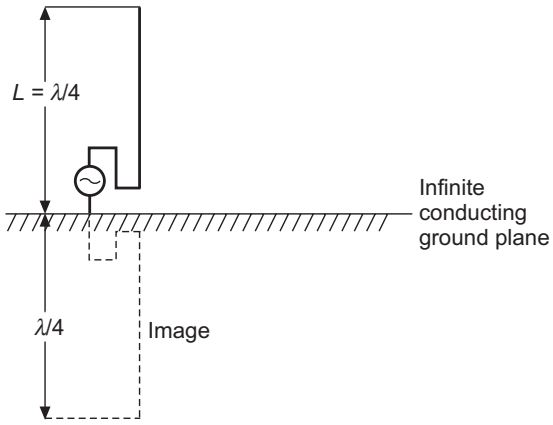


Figure 19.9(a) The monopole antenna. (Its image is also shown in the figure.)

Thus, in spite of the similarity, the following points should be borne in mind while analyzing a monopole antenna.

1. The radiation from a monopole is in the semi-infinite half-space as distinct from a dipole which radiates in the infinite space. Hence the range of θ in the monopole is from 0 to $\pi/2$ and for the dipole θ varies from 0 to π.
2. For identical currents in a monopole and the corresponding dipole, the radiated power by the monopole is only half of that radiated by the corresponding dipole.
3. Since the radiated power of a monopole is half of that radiated by the corresponding dipole for the same current, the input impedance of the monopole is half of the dipole.

19.4 THE MAGNETIC DIPOLE RADIATION

The static magnetic dipole has already been studied in Section 8.2.1., where we considered a circular current loop of radius a, and carrying a current I; and was found to have a vector potential $\mathbf{A}$ at a point P(r, ϕ, θ) from the centre of the dipole, as

$$\mathbf{A}_\phi = \mathbf{i}_\phi \left(\frac{\mu_0 I \pi a^2 \sin\theta}{4\pi r^2} \right) = \mathbf{i}_\phi \left(\frac{\mu_0 IS \sin\theta}{4\pi r^2} \right)$$

$$= \mathbf{i}_\phi \left(\frac{\mu_0 m \sin\theta}{4\pi r^2} \right) = \frac{\mu_0 \mathbf{m} \times \mathbf{u}}{4\pi r^2} \tag{8.11}$$

where the magnetic moment of the loop (or the dipole) was:

$$\mathbf{m} = \mathbf{i}_z IS, \text{ and } \mathbf{u} = \text{the unit vector in the direction of } \mathbf{r}$$

We now consider a magnetic dipole, as shown in Figure 19.10.

The vector potential at a point P(r, θ, ϕ) or (x, y, z) as shown in Figure 19.10, is the vector sum of the potential due to each element (Idl) of the dipole loop, i.e.

$$\mathbf{A} = \left(\frac{\mu_0 I}{4\pi} \right) \oint \frac{d\mathbf{l}}{r'} \tag{19.38}$$

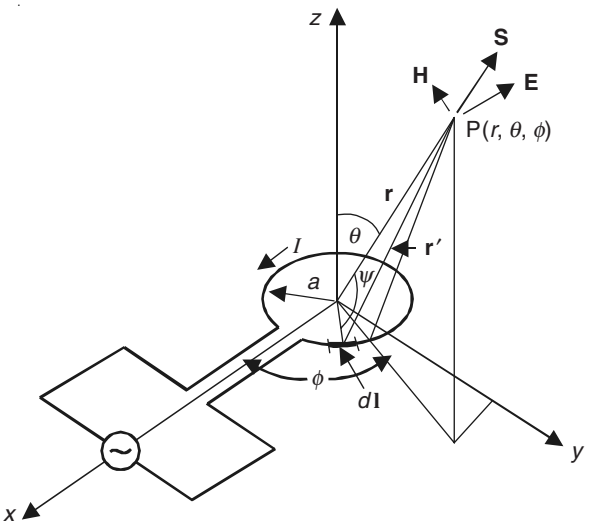


Figure 19.10 Magnetic dipole antenna fed by an oscillator.

where r' is the distance of the point P from the element under consideration. It can be easily seen that, in the loop, for any given value of r', there are two symmetrical $d\mathbf{l}$ vectors, such that their y-components add up and the x-components cancel.

$$\therefore \quad \mathbf{A} = \left[\left(\frac{\mu_0 I}{4\pi}\right)\int_0^{2\pi} \frac{a\, d\phi' \cos\phi'}{r}\right]\mathbf{i}_\phi \tag{19.39}$$

where ϕ' is the ϕ-coordinate of the element (Idl) under consideration; and ψ is the angle between the radius vector a of the current element (Idl) and r' (the distance between the point P and the element under consideration).

Now,
$$r'^2 = r^2 + a^2 - 2ra\cos\psi$$

$$\therefore \quad \frac{r}{r'} = \left[1 - \frac{a^2}{r'^2} + \frac{2ar}{r'^2}\cos\psi\right]^{1/2}$$

$$\simeq 1 - \frac{a^2}{2r'^2} + \frac{a}{r}\cos\psi$$

since $a << r'$, and $(r/r') \approx 1$, and hence we can substitute r for r' in the denominators of the two correction terms on the right-hand side. Now consider $(\mathbf{r}\cdot\mathbf{a})$, $\mathbf{a}$ being the vector radius at the element under consideration.

$$(\mathbf{r}\cdot\mathbf{a}) = (\mathbf{i}_x x + \mathbf{i}_y y + \mathbf{i}_z z)\cdot[(a\cos\phi')\mathbf{i}_x + (a\sin\phi')\mathbf{i}_y]$$

$$= xa\cos\phi' + ya\sin\phi' = ra\cos\psi$$

$$\therefore \quad \cos\psi = \left(\frac{x}{r}\right)\cos\phi' + \left(\frac{y}{r}\right)\sin\phi'$$

$$\therefore \quad \frac{r}{r'} = 1 - \frac{a^2}{2r^2} + \left(\frac{ax}{r^2}\right)\cos\phi' + \left(\frac{ay}{r^2}\right)\sin\phi' \tag{19.40}$$

Now, for the 'oscillating magnetic dipole', we have to consider the retardation effect, i.e.

$$A_P = \left(\frac{\mu_0}{4\pi}\right)\int_0^{2\pi}\left(\frac{1}{r'}\right)\left[I_0 \exp\left\{j\omega\left(t - \frac{r'}{c}\right)\right\}\right] a \cos\phi' \, d\phi' \tag{19.41}$$

$$= \frac{\mu_0 a I_0 \exp\left\{j\omega\left(t - \frac{r}{c}\right)\right\}}{4\pi} \int_0^{2\pi}\left(\frac{1}{r'}\right)\exp\left(\frac{j\omega(r - r')}{c}\right)\cos\phi' \, d\phi'$$

Since the loop is quite small compared with the wavelength, i.e.

$$|(r - r')| << \frac{\lambda}{2\pi}, \text{ then:}$$

$$\exp\left\{\frac{j\omega(r - r')}{\mathrm{c}}\right\} \simeq 1 + j\left(\frac{\omega}{\mathrm{c}}\right)(r - r')$$

$$\therefore \quad \mathbf{A} = \mathbf{i}_\phi\left[\frac{\mu_0 a I_0 \exp\left\{j\omega\left(t - \frac{r}{\mathrm{c}}\right)\right\}}{4\pi r}\right] \times \int_0^{2\pi}\left(\frac{r}{r'}\right)\left[1 + j\left(\frac{\omega}{\mathrm{c}}\right)(r - r')\right]\cos\phi' \, d\phi'$$

$$\text{The integral} = \int_0^{2\pi}\left[\frac{r}{r'} + j\left(\frac{\omega r}{\mathrm{c}}\right)\left(\frac{r}{r'} - 1\right)\right]\cos\phi' \, d\phi'$$

$$= \int_0^{2\pi}\left[\left\{1 - \frac{a^2}{2r^2} + \left(\frac{ax}{r^2}\right)\cos\phi' + \left(\frac{ay}{r^2}\right)\sin\phi'\right\}\right.$$

$$\left. + j\left(\frac{\omega r}{\mathrm{c}}\right)\left\{-\frac{a^2}{2r^2} + \left(\frac{ax}{r^2}\right)\cos\phi' + \left(\frac{ay}{r^2}\right)\sin\phi'\right\}\right]\cos\phi' \, d\phi'$$

$$= \left(\frac{ax}{r^2}\right)\pi + j\left(\frac{\omega r}{\mathrm{c}}\right)\left(\frac{ax}{r^2}\right)\pi$$

$$\therefore \quad \mathbf{A} = \mathbf{i}_\phi\left(\frac{\mu_0 I_0 \pi a^2}{4\pi r^2}\right)\exp\left\{j\omega\left(t - \frac{r}{\mathrm{c}}\right)\right\}\left\{1 + j\left(\frac{\omega r}{\mathrm{c}}\right)\right\}\sin\theta \tag{19.42}$$

where $x = r \sin \theta$.

$$\therefore \quad \mathbf{A} = \mathbf{i}_\phi\left[\frac{\mu_0 m_0 \exp\left\{j\omega\left(t - \frac{r}{\mathrm{c}}\right)\right\}}{4\pi r^2}\right]\left\{1 + j\left(\frac{\omega r}{\mathrm{c}}\right)\right\}\sin\theta$$

$$= \mathbf{i}_\phi\left[\frac{\mu_0[\mathbf{m}_0] \times \mathbf{r}_1}{4\pi r^2}\right]\left\{1 + j\left(\frac{\omega r}{\mathrm{c}}\right)\right\} \tag{19.43}$$

where m_0 is the magnetic moment of the dipole = $I_0 \pi a^2 = I_0 S$; and $[\mathbf{m}_0]$ is the retarded value of the moment; and $\mathbf{r}_1$ is the unit vector in the direction of r.

When $\omega = 0$, Eq. (19.43) reduces to the magnetic vector potential of the static dipole, i.e.

$$\mathbf{A} = \mathbf{i}_\phi r^2 \left[\left(\frac{\mu_0}{4\pi} \right) \left(\frac{\mathbf{m} \times \mathbf{r}_1}{\mathbf{r}^2} \right) \right] \tag{19.44}$$

Since there is no static charge in the dipole, $V = 0$. (19.45)

And, hence the Lorentz gauge simplifies to:

$$\text{div } \mathbf{A} = \nabla \cdot \mathbf{A} = 0 \tag{19.46}$$

19.4.1 E and H Vectors of the Oscillating Magnetic Dipole

$$\mathbf{H} = \left(\frac{1}{\mu_0} \right) \nabla \times \mathbf{A}$$

$$\therefore \quad H_r = \left(\frac{1}{\mu_0 r \sin\theta} \right) \left(\frac{\partial}{\partial \theta} \right) \left(A_\phi \sin\theta \right)$$

$$= \left[\frac{m_0 \exp\left\{ j\omega \left(t - \frac{r}{\mathrm{c}} \right) \right\}}{2\pi r^3} \right] \left[1 + j \left(\frac{\omega r}{\mathrm{c}} \right) \right] \cos\theta \tag{19.47}$$

$$H_\theta = -\left(\frac{1}{\mu_0 r} \right) \left(\frac{\partial}{\partial r} \right) (r A_\phi)$$

$$= -\left(\frac{m_0}{4\pi r} \right) \left(\frac{\partial}{\partial r} \right) \left[\exp\left\{ j\omega \left(t - \frac{r}{\mathrm{c}} \right) \right\} \left\{ \left(\frac{1}{r} \right) + j \left(\frac{\omega}{\mathrm{c}} \right) \right\} \right] \sin\theta$$

$$= -\left(\frac{m_0}{4\pi r} \right) \times \left[-j \left(\frac{\omega}{\mathrm{c}} \right) \left\{ \frac{1}{r} + j \left(\frac{\omega}{\mathrm{c}} \right) \right\} - \left(\frac{1}{r^2} \right) \right] \left[\exp\left\{ j\omega \left(t - \frac{r}{\mathrm{c}} \right) \right\} \right] \sin\theta$$

$$= \left[\frac{m_0 \exp\left\{ j\omega \left(t - \frac{r}{\mathrm{c}} \right) \right\}}{4\pi r^3} \right] \left[-\left(\frac{\omega^2 r^2}{\mathrm{c}^2} \right) + 1 + j \left(\frac{\omega r}{\mathrm{c}} \right) \right] \sin\theta \tag{19.48}$$

$$\mathbf{E} = -\left(\frac{\partial \mathbf{A}}{\partial t} \right) - \nabla V = -\left(\frac{\partial \mathbf{A}}{\partial t} \right)$$

$$= \mathbf{i}_\phi \left[\frac{\mu_0 m_0 \exp\left\{ j\omega \left(t - \frac{r}{\mathrm{c}} \right) \right\}}{4\pi r^3} \right] \left\{ j\omega - \left(\frac{\omega^2 r}{\mathrm{c}} \right) \right\} \sin\theta \tag{19.49}$$

Since the scalar potential is zero, **E** is hence due to the changing magnetic field. **H** lies in a plane passing through the z-axis, whereas **E** is azimuthal.

At $\omega = 0$, **H** reduces to that of a static magnetic dipole, and **E** = 0.

If we compare the **E** and the **H** fields of the magnetic and the electric dipoles, it is seen that the fields of the two are similar, except that the expressions for **E** and **H** are interchanged.

19.4.2 Poynting Vector and the Radiated Power of the Oscillating Magnetic Dipole

For $r >> \lambda$,

$$H_r = 0 \tag{19.50a}$$

$$H_\theta = -\left[\left(\frac{m_0\omega^2}{4\pi rc^2}\right)\sin\theta\right]\exp\left\{j\omega\left(t-\frac{r}{\mathrm{c}}\right)\right\} \tag{19.50b}$$

$$E_\phi = \left[\left(\frac{\mu_0 m_0\omega^2}{4\pi rc}\right)\sin\theta\right]\exp\left\{j\omega\left(t-\frac{r}{\mathrm{c}}\right)\right\} \tag{19.50c}$$

$$\therefore \quad \bar{\mathbf{S}}_{\mathrm{av}} = \mathbf{i}_r\left(\frac{\mu_0\omega^4 m_0^2}{32\pi^2 r^2\mathrm{c}^3}\right)\sin^2\theta \tag{19.51}$$

The total radiated power W is obtained by integrating $\mathbf{S}_{\mathrm{av}}$ over the surface of a sphere of radius r, i.e.

$$W = \frac{\mu_0\omega^4 m_0^2}{32\pi^2\mathrm{c}^3}\int_0^{2\pi}\int_0^{\pi/2}\left(\frac{\sin\theta}{r^2}\right)r^2\sin\theta\,d\theta\,d\phi$$

$$= \left(\frac{\mu_0\omega^4 m_0^2}{32\pi^2\mathrm{c}^3}\right)(2\pi)\left(\frac{4}{3}\right) = \frac{\mu_0\omega^4 m_0^2}{12\pi\mathrm{c}^3}$$

Substituting $\frac{\omega}{\mathrm{c}} = \frac{2\pi}{\lambda}$, and $m_0 = I_0\pi a^2$,

$$W = \left(\frac{\mu_0\mathrm{c}}{12\pi}\right)\left(\frac{2\pi}{\lambda}\right)^4(I_0\pi a^2)^2 = \left(\frac{\mu_0\mathrm{c}\pi}{12}\right)\left(\frac{2\pi a}{\lambda}\right)^4 I_0^2 \tag{19.52}$$

The radiation resistance is the coefficient of $\left(\frac{I_0^2}{2}\right)$.

$$\therefore \quad R_{\mathrm{rad}} = 197\left(\frac{2\pi a}{\lambda}\right)^4\Omega, \quad a << \frac{\lambda}{2\pi} \tag{19.53}$$

Thus the radiation resistance of the magnetic dipole is proportional to the 4th power of the frequency, whereas that of the electric dipole is proportional to the square of the frequency. This dipole is the basis for loop antenna.

It will be noticed that E_ϕ given by Eq. (19.50) is a maximum, when $\theta = (\pi/2)$, i.e. in the plane of the loop. Hence, by the reciprocity theorem (which we shall study later in this chapter), the signal received by a loop will be a maximum when the plane of the loop passes through the transmitter. This property is used in the loop antenna which is fitted to the aircraft and to the ships to locate their positions by taking bearings on the radio beacons at known places.

19.4.3 Electric and Magnetic Quadrupoles

We now go on to study the more elaborate types of radiation produced by quadrupoles. We first study the linear electric quadrupole which consists of two dipoles of opposite polarity arranged in a line to give three charges $+q$, $-2q$, $+q$, at distances l from each other. The dipole moment of such charge distribution is zero (i.e. $m = \Sigma (Q_\mathrm{m}\, l) = 0$, where $q = f(t) = Q_\mathrm{m} \sin \omega t$, Q_m being the amplitude of the charge). Its quadrupole moment is $m_{00} = \Sigma (Q_\mathrm{m}\, l^2) = 2Q_\mathrm{m}\, l^2$.

Since the dipole moment is zero, there will be no dipole radiation, but as there are time-varying charges, there will be radiation of some sort which would be the subject of our investigation. We can proceed exactly as for the electric dipole radiation, and calculate as before the quantities V, $\mathbf{A}$, $\mathbf{E}$, and $\mathbf{H}$. However, it is easier to add the fields of the component dipoles as indicated below; and for simplicity we consider the radiation field components at large distances ($r >> \lambda$, i.e. the far field). There are two dipoles now, one with the moment $[-m_0 \exp(j\omega t)]$ centred at $(-l/2)$, and another with the moment $[+m_0 \exp(j\omega t)]$ centred at $(+l/2)$. The electric field intensities of the two dipoles can be added vectorially (by the principle of superposition). Since the two dipoles (forming the quadrupole) are located at some distance away from the origin of the coordinate system, their field intensities at a point P(r, θ, ϕ) will differ slightly in direction, in amplitude, and in phase. However for the far field (when $r >> l$), the differences in direction and in amplitude may easily be neglected, but not in phase. Hence, for each dipole at the origin,

$$\mathbf{E}_\theta = \mathbf{i}_\theta \left[-\left(\frac{m_0 \omega^2}{4\pi\varepsilon_0 \mathrm{c}^2 r} \right) \sin\theta \exp\left\{ j\omega \left(t - \frac{r}{\mathrm{c}} \right) \right\} \right] \qquad \text{[from Eq. (19.15a)]}$$

For the quadrupole,

$$\mathbf{E}_\theta = \mathbf{i}_\theta \left[-\left(\frac{m_0 \omega^2}{4\pi\varepsilon_0 \mathrm{c}^2 r} \right) \sin\theta \exp\left\{ j\omega \left(t - \frac{r}{\mathrm{c}} \right) \right\} \right] \times \left[\exp\left\{ j\left(\frac{\omega l}{2\mathrm{c}} \right) \right\} \cos\theta - \exp\left\{ -j\left(\frac{\omega l}{2\mathrm{c}} \right) \right\} \cos\theta \right]$$

These two exponential functions can be expanded in power series in $(\omega l/\mathrm{c})$ for the far field, and then their sum reduces to $j(\omega l/\mathrm{c}) \cos\theta$.

$$\therefore \quad \mathbf{E}_\theta = \mathbf{i}_\theta \left[-j \left(\frac{m_0 \omega^3}{8\pi\varepsilon_0 \mathrm{c}^3 r} \right) \sin\theta \cos\theta \exp\left\{ j\omega \left(t - \frac{r}{\mathrm{c}} \right) \right\} \right] \quad \text{for } r >> \lambda >> l \qquad (19.15\mathrm{a}')$$

There can be no radiation along the axis $\theta = 0$ or $= \pi$, where neither of the two dipoles forming the quadrupole radiate energy. Also along the equator $\theta = \pi/2$, there would be zero radiation as the two dipoles produce equal and opposite fields on this plane.

The magnetic field intensity can be found directly from the $\mathbf{E}$ field as

$$\mathbf{H}_\phi = \mathbf{i}_\phi\, E_\theta \left(\frac{\mu_0}{\varepsilon_0} \right)^{-1/2}$$

Next, we consider an oscillating magnetic quadrupole which can be formed from two oscillating magnetic dipole loops placed parallel to each other on either side of the origin of the coordinate system. For the simplest case, we assume the distance between the two loops to be equal to the radius (= a) of the loop. The lower dipole is centred at $z = -a/2$ and has a moment, $-m_0 \exp(j\omega t)$, whereas the upper dipole is located at $z = +a/2$ and has a moment, $+m_0 \exp(j\omega t)$. This arrangement is very similar to that of the linear electric quadrupole discussed in the earlier part of this section.

Once again, the fields of the two magnetic dipoles can be added up, neglecting the differences in direction and amplitude, but taking into account the difference in phase or retardation. This is done by multiplying the dipole field by $j(a\omega/2\pi c)\cos\theta$,

$$H_\theta = -\left[j\left(\frac{m_0 a\omega^3}{8\pi^2 r c^3} \right) \sin\theta\cos\theta \right] \exp\left\{ j\omega\left(t - \frac{r}{c} \right) \right\} \qquad (19.50b')$$

$$E_\phi = \left[j\left(\frac{\mu_0 m_0 a\omega^3}{8\pi^2 r c^2} \right) \sin\theta\cos\theta \right] \exp\left\{ j\omega\left(t - \frac{r}{c} \right) \right\} \quad \text{for } r >> \lambda >> a \qquad (19.50c')$$

19.4.4 Antenna Characteristics (Radiation Parameters)

So far we have considered the theory of some of the basic elementary antenna types. Now we shall discuss some of the important characteristics of an antenna as radiator of electromagnetic energy. These are stated as below.

(A) An Antenna Pattern (or Radiation Pattern)

This is a three-dimensional plot of its radiation at far field. So this parameter describes the directional dependence of the radiated power of the antenna.

The plot of the amplitude of a specified component of the E-field is called its field pattern, the plot of square of the amplitude of E is its power pattern. A three-dimensional plot can be avoided by plotting separately the normalized $|E|$ against θ for a given ϕ (in spherical polar coordinate system). This is called E-plane pattern or vertical pattern. When $\phi = \pi \times 2$, the pattern is called H-plane pattern or horizontal pattern. The normalization of E is done with respect to maximum value of E.

For Hertzian dipole, from Eqs. (19.14a) and (19.14b), it is seen that

$$f(\theta) = \sin\theta \qquad (19.53a)$$

and the polar or angular plot will be as shown in Figures 19.10(a), (b) and (c).

In general, a radiation pattern has a direction of maximum radiation as shown in the polar plot of Figure 19.10(d). This is the main lobe. There are certain directions along which there is no radiation. Such directions are called nulls. There are also certain directions along which the radiation has a local maximum and these are called 'side lobes'.

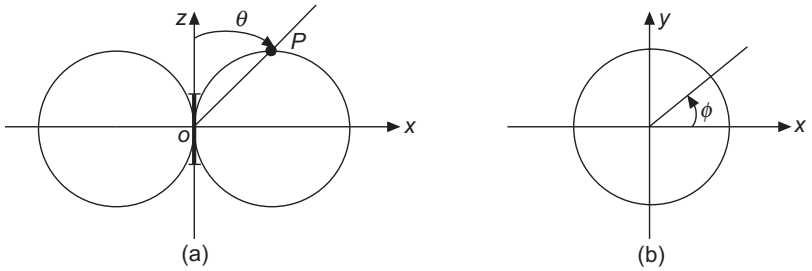


Figure 19.10(a) and (b) Field patterns of the Hertzian dipole: (a) normalized *E*-plane, $\phi = 0$; (b) normalized *H*-plane, $\theta = \pi/2$.

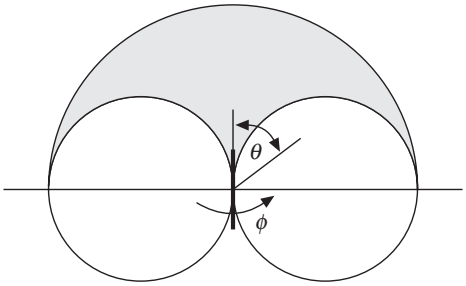


Figure 19.10(c) Three-dimensional pattern.

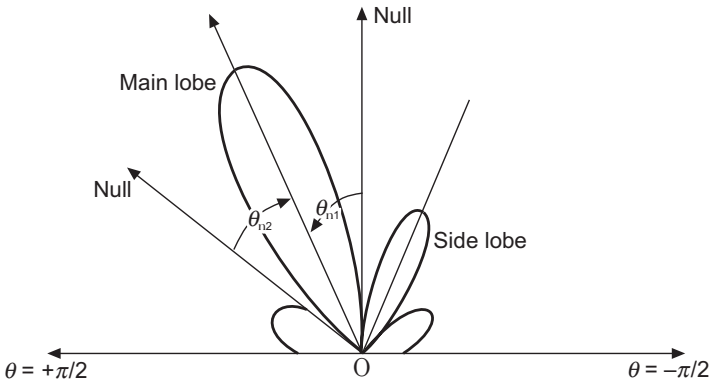


Figure 19.10(d) Radiation pattern as a polar plot.

Direction of the main beam: This is the direction along which the radiation field strength is maximum, and is denoted by Q_{max}.

Half-power beam width (HPBW): The angular region, where the effective radiation from the antenna goes, is the main beam.

The effective width of the main beam = the angular width of the pattern between the points on the radiation pattern where the magnitude of the field reduces to $\frac{|E_{max}|}{\sqrt{2}}$, E_{max} being the maximum field.

From Figure 19.10(e),

$$\text{HPBW} = \theta_1 - \theta_2 \tag{19.53b}$$

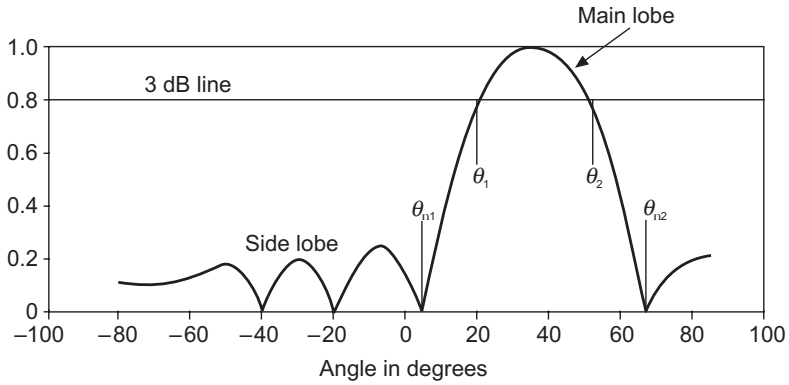


Figure 19.10(e) Cartesian plot of the radiation pattern.

Since the field in the directions of θ_1 and θ_2 reduces by a factor of $\frac{1}{\sqrt{2}}$ of $E_{\max}$, the Poynting vector in these directions reduces to 1/2 or –3 dB compared to its value in the direction of maximum radiation. (HPBW is also referred to as '3 dB beam width' of the antenna). HPBWs are generally measured in the planes of E and H.

Beam width between first nulls (BWFN): Sometimes, the width of the main beam is measured by the angular separation between the first nulls on either side of the direction of the main lobe.

$$\therefore \qquad \text{BWFN} = \theta_{n2} - \theta_{n1} \tag{19.53c}$$

However HPBW is a better measure of the effective width of the main beam as the shape of the main beam can change while keeping the null positions unchanged.

Side lobe level (SLL): Side lobes indicate the leakage of power in the undesired directions. Ideally there should be no radiation outside the main beam. But in any practical antenna, there are side lobes and the total radiated power is not focussed into the main beam. So this leakage is to be kept minimal, i.e. the side lobe amplitude should be kept as small as possible compared to the main beam amplitude. For good satellite antennae, the SLL is generally –30 dB to –40 dB. When we move away from the main beam, the amplitudes of the side lobes decrease and hence the first side lobe usually defines the SLL.

(B) Radiation Intensity

We know that the radiation pattern (of an antenna) is a three-dimensional surface and the radiation intensity is, in general, a function of θ and ϕ and is given by

$$U(\theta, \phi) = \frac{\text{Power along the direction } (\theta, \phi) \text{ in a solid angle } d\Omega}{\text{Solid angle } (d\Omega)} \tag{19.53d}$$

The solid angle on the surface of a sphere,

$$d\Omega = \frac{dA}{r^2} \tag{19.53e}$$

where

dA = area of the surface of the sphere

r = radius of the sphere

$d\Omega = \sin\theta\, d\theta\, d\phi$, i.e. differential solid angle

$$\therefore \quad \left.\begin{aligned} U(\theta, \phi) &= \frac{\text{Power along } (\theta, \phi)}{dA} r^2 \\ &= (\text{Power density}) r^2 \end{aligned}\right\} \tag{19.53f}$$

Now, power density of a radiated wave = magnitude of the Poynting vector

i.e. $$P(\theta, \phi) = \frac{|E(\theta, \phi)|^2}{\eta} \tag{19.53g}$$

(η = intrinsic impedance of the medium)

$$\therefore \quad U(\theta, \phi) = \frac{|E(\theta, \phi)|^2}{\eta} r^2 = P(\theta, \phi) r^2 \tag{19.53h}$$

The total power radiated (= P_{rad}) can be obtained by integrating the radiation intensity over the total solid angle (= 4π), i.e.

$$P_{\text{rad}} = \iint_{\Omega = 4\pi} U(\theta, \phi)\, d\Omega = \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} U(\theta, \phi) \sin\theta\, d\theta\, d\phi \tag{19.53i}$$

Hence the average radiation intensity

$$U_{\text{av}} = \frac{P_{\text{rad}}}{4\pi} = \frac{1}{4\pi} \iint U(\theta, \phi)\, d\Omega \tag{19.53j}$$

(C) Directive Gain and Directivity

The radiation focussing capability of an antenna is quantified by the parameter 'directivity'. So as a first step in this process we define a parameter called its 'directive gain'. So, the **Directive Gain** G_D (θ, ϕ) of an antenna is a measure of the concentration of the radiated power in specified direction (θ, ϕ). It is thus the measure of the ability of the antenna to direct the radiated power in a given direction.

So it is the ratio of the radiation intensity in a given direction (θ, ϕ) to the average radiation intensity, i.e.

$$G_D(\theta, \phi) = \frac{U(\theta, \phi)}{U_{\text{av}}}$$

$$= \frac{4\pi\, U(\theta, \phi)}{\iint U(\theta, \phi)\, d\Omega} = \frac{4\pi\, U(\theta, \phi)}{P_{\text{rad}}} \tag{19.53k}$$

So $P(\theta, \phi)$ can be expressed in terms of the directive gain as

$$P(\theta, \phi) = \frac{G_D}{4\pi r^2} P_{\text{rad}} \tag{19.53l}$$

The directive gain $G_D(\theta, \phi)$ depends on antenna pattern. For Hertzian dipole and also for $\lambda/2$ dipole, $P(\theta, \phi)$ is maximum for $\theta = \pi/2$, and minimum (i.e. = 0) at $\theta = 0$ or π. So the Hertzian dipole radiates power in a direction broadside its length.

In general G_D can vary from 0 to ∞. In the direction of the nulls, $E(\theta, \phi)$ and hence $U(\theta, \phi)$ and $G(\theta, \phi)$ are zero, whereas G_D is maximum in the direction of the main beam.

So now we define directivity:

The directivity D of an antenna is the ratio of the maximum radiation intensity to the average radiation intensity.

Thus the directivity D is the maximum directive gain.

$$\therefore \quad D = \frac{U_{\max}}{U_{\text{av}}} = G_{D\max}$$

or

$$D = \frac{4\pi U_{\max}}{P_{\text{rad}}} = \frac{4\pi U_{\max}}{\iint U(\theta, \phi)\, d\Omega} \tag{19.53m}$$

Thus $D = 1$ for an isotropic antenna (which does not exist physically and is an ideal reference).

So,

$$D = \frac{4\pi E_{\max}^2}{\int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} \left|E_{\theta, \phi}\right|^2 \sin\theta\, d\theta\, d\phi} \tag{19.53n}$$

Namalizing the radiation pattern as $E_n(\theta, \phi) = \dfrac{E(\theta, \phi)}{E_{\max}}$, D can be rewritten as

$$D = \frac{4\pi}{\int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} \left|E_n(\theta, \phi)\right| \sin\theta\, d\theta\, d\phi} \tag{19.53p}$$

For a Hertzian dipole, $G_D(\theta, \phi) = 1.5 \sin^2\theta, \quad D = 1.5$

For a $\lambda/2$ dipole, $G_D(\theta, \phi) = \dfrac{\eta}{\pi R_{\text{rad}}} f^2(\theta), \quad D = 1.64$ (19.53q)

where $\eta = 120\pi$, $R_{\text{rad}} = 73\ \Omega$, and $f(\theta) = \dfrac{\cos\left(\frac{\pi}{2}\cos\theta\right)}{\sin\theta}$

(D) Antenna Gain or Power Gain

As seen above, the parameter directivity depends totally on the radiation pattern. An implicit assumption in this discussion, has been that the total power radiated P_{rad}, is same as the power

supplied to the antenna input. In actual practice, the antennae are made out of conductors having finite resistivity (or conductivity), and so when there is a current flow in the antenna surface, there is an ohmic loss. So a part of the power supplied to the antenna input is lost in the heating of the antenna due to the ohmic loss. So if P_i is the power supplied to the antenna input, then

$$P_i = P_{\text{rad}} + P_l \tag{19.53r}$$

where P_l is the ohmic loss in the antenna due to its finite conductivity.

$$\therefore \qquad P_i = \frac{1}{2} |I_{\text{in}}|^2 (R_l + R_{\text{rad}}) \tag{19.53s}$$

where

I_{in} = current at the input terminals

R_l = loss or ohmic resistance of the antenna.

So we define the power efficiency (= η_r) of the antenna as

$$\eta = \frac{P_{\text{rad}}}{P_i} = \frac{P_i - P_l}{P_i} = \frac{P_{\text{rad}}}{P_{\text{rad}} + P_l} \tag{19.53t}$$

And the antenna power gain (= G_P) is defined as

$$\left.\begin{aligned} G_P(\theta, \phi) &= \frac{U_{\max}(\text{actual})}{U_{\text{av}} \text{ for a loss-less case}} \\ &= \frac{4\pi U_{\max} \text{ (actual)}}{P_i} \\ &= \frac{4\pi U_{\max}(\text{actual})}{P_{\text{rad}}} \cdot \frac{P_{\text{rad}}}{P_i} = D\eta_r \end{aligned}\right\} \tag{19.53u}$$

Thus

$$\text{Power Gain} = \text{Directivity} \times \text{Efficiency} \tag{19.53v}$$

19.5 ANTENNA ARRAYS

19.5.1 Directional Properties of Antennae

The antennae have two main functions. The first job is to radiate the high frequency energy which has been generated in the transmitter, and then guided to the antenna by the transmission line. In this process, the antenna is acting as an impedance matching device whereby it matches the impedance of the transmission line to that of the free space. The second job of the antenna is to direct the energy to the required directions and to suppress the energy from those directions where it is not required.

In general, a completely omni-directional (i.e. non-directional) radiator radiates energy uniformly in all directions. Such a device is called an isotropic radiator or a unipole, an example of this being a point source of sound. Since all radio antennae have some directivity, there is no isotropic radiator of electromagnetic energy. However, conceptually such a hypothetical source can still be used as a standard of reference for comparing with actual devices.

The radiation pattern of an antenna is a graphical representation of the radiation of the antenna as a function of the direction. The radiation can be expressed either in terms of the field-

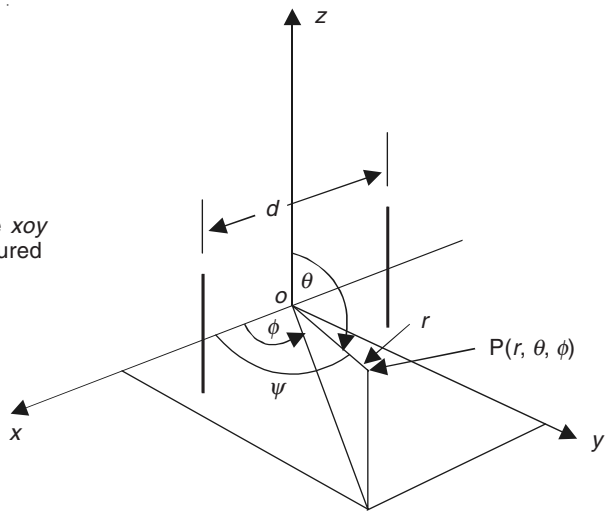
strength (i.e. **E** in volts/metre) or in terms of the power per unit solid angle. The power pattern is proportional to the square of the field strength pattern. The coordinate system, generally used for specifying the radiation patterns is the spherical polar coordinate system (r, θ, ϕ). The antenna is usually located at or near the origin of the coordinate system, and the field strength is specified at points on the spherical surface of radius r. So long as r is chosen such that $r >> \lambda$ and also >> the largest dimension of the antenna system, the shape of the radiation pattern is independent of r. For the radiation field, **E** is always tangential to the spherical surface. For a vertical dipole, **E** is in the θ-direction; and for a horizontal loop, **E** is in the ϕ-direction. In general, there may be both E_θ and E_ϕ components which may or may not be in time-phase. In this case, the characteristics are shown by separate patterns for θ- and ϕ-polarizations, instead of the older vertical and horizontal polarizations respectively. The older denominations were somewhat confusing, because the θ- or the vertical polarization does not rigorously mean that the signal is vertical (though the signal does lie in a vertical plane through the radius vector), whilst the ϕ- or the horizontally polarized signal is always horizontal.

The half-wave wire antenna which we have studied so far, is directive, in the sense that most of its radiation is in the direction for which $\theta > 60°$, but it does not discriminate between the azimuthal directions. To produce a preferred direction in the equatorial plane, it is necessary to replace the single antenna by a group or an array of antennae suitably spaced and excited. So we define an 'antenna array' as a system of similar antennae, similarly oriented. The arrays make use of the wave-interference phenomena which take place between the radiations from the different elements of the array. We shall consider first, the simplest two-element array.

19.5.2 Two Element Array

We consider the simplest array in which two half-wave antennae (with their axes parallel to the z-axis) are spaced apart $\lambda/2$, and are excited in phase. The antennae are parallel and lie in the x-z plane, as shown in Figure 19.11. The point of interest P is (r, θ, ϕ) such that $r >> \lambda$.

The **E** field due to each antenna is given by Eq. (19.27) which we can rewrite as



Note: ϕ is measured in the *xoy* plane whereas ψ is measured in the *xo*P plane.

Figure 19.11 A pair of parallel half-wave antennae, d metres apart.

$$E_\theta = \left[\frac{j60I_0 \exp\left\{j\omega\left(t - \frac{r}{\mathrm{c}}\right)\right\}}{r}\right]\left[\frac{\cos\left\{\left(\frac{\pi}{2}\right)\cos\theta\right\}}{\sin\theta}\right] \tag{19.54}$$

For the array, the resulting E_θ will be

$$E_\theta = j\left(\frac{60I_0}{r}\right)\left[\frac{\cos\left\{\left(\frac{\pi}{2}\right)\cos\theta\right\}}{\sin\theta}\right]\times\left[\exp\left\{j\omega\left(t - \frac{r + \frac{d\cos\psi}{2}}{\mathrm{c}}\right)\right\}\right.$$

$$\left. + \exp\left\{j\omega\left(t - \frac{r - \frac{d\cos\psi}{2}}{\mathrm{c}}\right)\right\}\right]$$

$$= \left[\frac{j60I_0}{r}\exp\left\{j\omega\left(t - \frac{r}{\mathrm{c}}\right)\right\}\right]\left[\frac{\cos\left\{\left(\frac{\pi}{2}\right)\cos\theta\right\}}{\sin\theta}\right]$$

$$\times\left[\exp\left\{-j\left(\frac{\omega}{\mathrm{c}}\right)\left(\frac{d\cos\psi}{2}\right)\right\} + \exp\left\{j\left(\frac{\omega}{\mathrm{c}}\right)\left(\frac{d\cos\psi}{2}\right)\right\}\right]$$

$$= \left[\frac{j60I_0}{r}\exp\left\{j\omega\left(t - \frac{r}{\mathrm{c}}\right)\right\}\right]\times\left[\frac{\cos\left\{\left(\frac{\pi}{2}\right)\cos\theta\right\}}{\sin\theta}\right]\cos\left[\left(\frac{\omega d}{2\mathrm{c}}\right)\cos\psi\right] \tag{19.55}$$

It should be noted that in the above expressions, of the two waves reaching the point P from the two antennae, one has travelled a distance $[r + (d\cos\psi)/2]$, and the other one a distance of $[r - (d\cos\psi)/2]$, and as a result, when they reach P, they will be out of phase. This, of course, assumes that the antennae had been excited in phase. Now, we replace the angle ψ by θ and ϕ, because:

$$r\cos\psi = r\sin\theta\cos\phi \tag{19.56}$$

$\therefore$ The electric field intensity expression becomes

$$\mathbf{E}_\theta = \mathbf{i}_\theta\left[\frac{j120I_0}{r}\exp\left\{j\omega\left(t - \frac{r}{\mathrm{c}}\right)\right\}\right]\times\left[\frac{\cos\left\{\left(\frac{\pi}{2}\right)\cos\theta\right\}}{\sin\theta}\right]\cos\left[\left(\frac{\omega d}{2\mathrm{c}}\right)\sin\theta\cos\phi\right] \tag{19.57}$$

When the two antennae are half-wavelength apart,

$$\frac{\omega d}{2\mathrm{c}} = \frac{\omega\lambda}{4\mathrm{c}} = \left(\frac{2\pi}{\lambda}\right)\left(\frac{\lambda}{4}\right) = \frac{\pi}{2}$$

$\therefore$ In the x-y plane, where $\theta = \pi/2$, the variation of the **E** field is proportional to $\cos\left(\frac{\pi}{2}\cos\theta\right)$, and hence

$$\mathbf{E} = \mathbf{i}_\theta \left[\frac{j120 I_0}{r} \exp\left\{j\omega\left(t - \frac{r}{c}\right)\right\}\right]\left[\cos\left\{\left(\frac{\pi}{2}\right)\cos\phi\right\}\right] \tag{19.58}$$

At points where $\phi = 0$ or π, $\mathbf{E} = 0$; and where $\phi = \pi/2$, **E** is a maximum, i.e.

$$\mathbf{E}_{max} = \mathbf{i}_\theta \left[\frac{j120 I_0}{r} \exp\left\{j\omega\left(t - \frac{r}{c}\right)\right\}\right] \tag{19.59}$$

Thus there is constructive interference along the y-axis; and destructive interference along the x-axis.

In the x-z plane, $\phi = 0$, and hence **E** is

$$\mathbf{E} = \mathbf{i}_\theta \left[\frac{j120 I_0}{r} \exp\left\{j\omega\left(t - \frac{r}{c}\right)\right\}\right] \times \left[\frac{\cos\left\{\left(\frac{\pi}{2}\right)\cos\theta\right\}}{\sin\theta}\right]\left[\cos\left\{\left(\frac{\pi}{2}\right)\sin\theta\right\}\right] \tag{19.60}$$

The first trigonometric term is the angular distribution for a single half-wave antenna. This is zero at $\theta = 0$, and maximum at $\theta = \pi/2$. The second trigonometric term is a consequence of the interference between the two antennae. It is a maximum at $\theta = 0$, and zero at $\theta = \pi/2$.

$\therefore$ In the x-z plane, **E** is zero both at $\theta = 0$, and $\theta = \pi/2$.

Finally, we consider the y-z plane where $\phi = \pi/2$.

$$\mathbf{E} = \mathbf{i}_\theta \left[\frac{j120 I_0}{r} \exp\left\{j\omega\left(t - \frac{r}{c}\right)\right\}\right] \times \left[\frac{\cos\left\{\left(\frac{\pi}{2}\right)\cos\theta\right\}}{\sin\theta}\right] \tag{19.61}$$

i.e. **E** on the plane y-z, behaves like the **E** field of a single half-wave antenna.

To obtain the expression for power, we rewrite the expression for the **E** vector, changing its numerical constant to the characteristic impedance, and we have

$$\mathbf{E} = \mathbf{i}_\theta \left[j\left(\frac{I_0 Z_0}{\pi r}\right)\exp\left\{j\omega\left(t - \frac{r}{c}\right)\right\}\right] \times \left[\frac{\cos\left\{\left(\frac{\pi}{2}\right)\cos\theta\right\}}{\sin\theta}\right]\left[\cos\left\{\left(\frac{\omega d}{2c}\right)\sin\theta\cos\phi\right\}\right] \tag{19.62a}$$

$$\mathbf{H} = \mathbf{i}_\phi \left[j\left(\frac{I_0}{\pi r}\right)\exp\left\{j\omega\left(t - \frac{r}{c}\right)\right\}\right] \times \left[\frac{\cos\left\{\left(\frac{\pi}{2}\right)\cos\theta\right\}}{\sin\theta}\right]\left[\cos\left\{\left(\frac{\omega d}{2c}\right)\sin\theta\cos\phi\right\}\right] \tag{19.62b}$$

$\therefore$ The average value in time, of the Poynting vector is

$$\bar{S}_{av} = \left(\frac{Z_0 I_0^2}{2\pi^2 r^2}\right) \times \left[\frac{\cos\left\{\left(\frac{\pi}{2}\right)\cos\theta\right\}}{\sin\theta}\right]^2 \cos^2\left[\left(\frac{\omega d}{2c}\right)\sin\theta\cos\phi\right] \tag{19.63}$$

This is for the two half-wave antennae spaced a distance d on the x-axis, both excited with the same current and in phase (by the same current, we mean both have currents of the same amplitude). We can generalize further, and if there is a phase difference between the antennae of α, then, we have:

$$\bar{S}_{av} = \left(\frac{Z_0 I_0^2}{2\pi^2 r^2}\right) \times \left[\frac{\cos\left\{\left(\frac{\pi}{2}\right)\cos\theta\right\}}{\sin\theta}\right]^2 \cos^2\left[\left(\frac{\omega d}{2c}\right)\sin\theta\cos\phi + \alpha\right] \tag{19.64}$$

This is the 'power per unit solid angle', when $r = 1$.

If the two antennae are in phase, and spaced a distance of half-wavelength, then $\alpha = 0$, $d = \lambda/2$, and $\{\omega d/(2c)\} = \pi/2$, and so:

$$\bar{S}_{av} = \left(\frac{Z_0 I_0^2}{2\pi^2 r^2}\right) \times \left[\frac{\cos\left\{\left(\frac{\pi}{2}\right)\cos\theta\right\}}{\sin\theta}\right]^2 \cos^2\left[\left(\frac{\pi}{2}\right)\sin\theta\cos\phi\right] \tag{19.65}$$

The pattern in the plane $\theta = \pi/2$, is shown in Figure 19.12, which indicates that there are two main directions of radiation which are both perpendicular to the line of the array. This is a particular example of a 'broadside array'. If the number of elements is increased, the main beam becomes sharper and small sidelobes appear.

On the other hand, if the two antennae are spaced quarter wavelength, i.e. $d = \lambda/4$, with a phase difference of $\pi/2$, i.e. $\alpha = \pi/2$, then Eq. (19.64) becomes

$$\bar{S}_{av} = \left(\frac{Z_0 I_0^2}{2\pi^2 r^2}\right) \times \left[\frac{\cos\left\{\left(\frac{\pi}{2}\right)\cos\theta\right\}}{\sin\theta}\right]^2 \cos^2\left[\left(\frac{\pi}{4}\right)(\sin\theta\cos\phi + 1)\right] \tag{19.66}$$

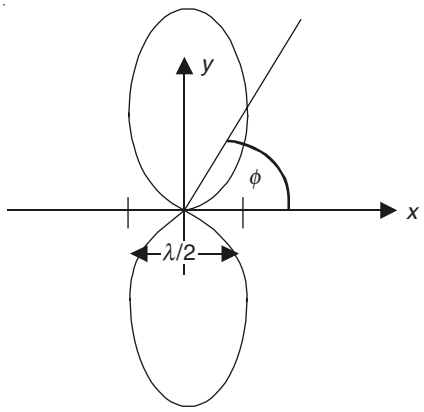


Figure 19.12 Two-element broadside array.

The pattern of this expression is symmetrical about the x-z plane (i.e. **E** plane) and about the x-y plane (i.e. **H** plane). The power patterns in these two planes are shown in Figure 19.13. It will be seen that only a small portion of the power is radiated to the left of $x = 0$, the main energy going in a broad beam in the direction of the positive x-axis, i.e. the line of the array. This system is an example of an 'end-fire array'.

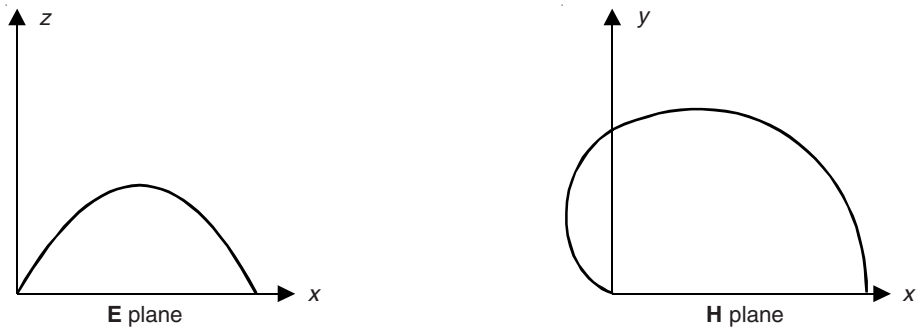


Figure 19.13 Two-element end-fire array.

As mentioned earlier, the factor $[\cos\{(\pi/2)\cos\theta\}/\sin\theta]^2$ in Eqs. (19.65) and (19.66) represents the pattern due to an element of the array. The factors $\cos^2[(\pi/2)\cos\theta]$ and $\cos^2[(\pi/4)(\sin\theta\cos\phi + 1)]$ characterize the broadside and the end-fire arrays respectively, and are called 'array factors'. The broadside array factor is like a toroid with a directive pattern in any plane through the x-axis. The end-fire array has a pattern like a pencil beam possessing directivity in both the planes through x- and y-axes. These arguments hold for the arrays containing more than two elements. The end-fire arrays usually have a higher gain than the broadside arrays of the same size. More highly directive systems can be designed by combining the two types which we have discussed now.

19.5.2.1 Two element array (of isotropic antennae)

So far we have described a two-element array of half-wave dipoles which is a practical one. Now we will discuss a two-element array made up of isotropic antennae. An isotropic antenna does not exist in real life and is a hypothetical one. But its use helps us in understanding the principle of the arrays, and we will show that an array of non-isotropic elements can be derived very easily from this analysis of isotropic element array.

Since we are considering an array of isotropic elements, this means that each antenna is a 'point source'. Let the two such antennae be separated by a distance d, and be excited by the currents $I_1 \angle\delta_1$ and $I_2 \angle\delta_2$ where I_1 and I_2, and δ_1, δ_2 are the amplitudes and phases of the excitation current in the two antennae respectively. The line joining the two elements is the axis of the array [i.e. line AB in Figure 19.13(a)]. All the angles measured from the axis of the array are ϕ which should not be confused with the spherical polar coordinate variable (r, θ, ϕ). As the antenna elements are point sources, the radiation pattern of the array would be axi-symmetric about the axis AB. So the section of the radiation pattern in a plane passing through this axis [e.g. (say) the plane of the paper] would generate the total pattern by rotating the section about AB. Since we are interested in the far field, let P be such a point at which the total **E** field has to be evaluated. The distance of the point P from the array is specified as r, as shown in Figure 19.13(a). This distance r is >> than the dimensions of the array.

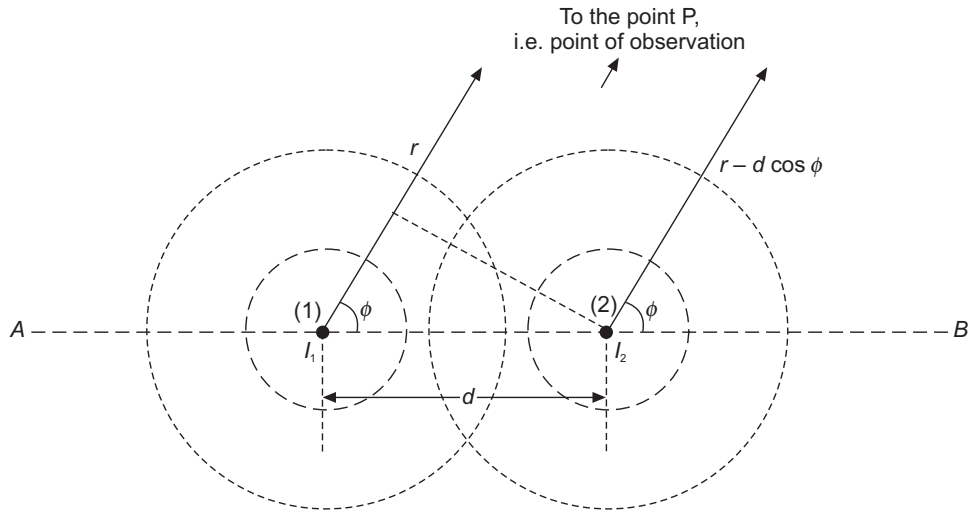


Figure 19.13(a) A two-element array of isotropic antennae.

∴ **E** field at the point P, due to the antenna (1) is

$$E_1 = \frac{C\, I_1 e^{j\delta_1}}{r}\, e^{-j\beta r} \tag{19.66a}$$

where C is a constant dependent on the parameters of the medium. Similarly the field E_2 at P due to the antenna (2) is

$$E_2 = \frac{C\, I_2 e^{j\delta_2}}{(r - d\cos\phi)}\, e^{-j\beta(r - d\cos\phi)} \tag{19.66b}$$

Since $d << r$, it can be neglected in the denominator of the amplitude term, but not in the phase term.

$$\therefore \quad E_2 = \frac{C\, I_2\, e^{j\delta_2}}{r}\, e^{-j\beta(r - d\cos\phi)} \tag{19.66c}$$

Since the two isotropic antennae are identical, the directions of E_1 and E_2 will be same.

$$\therefore \quad E_T = E_1 + E_2$$

$$= \frac{C\, e^{-j\beta r}}{r}\left(I_1 e^{j\delta_1} + I_2 e^{j\delta_2}\, e^{j\frac{2\pi d}{\lambda}\cos\phi} \right)$$

For a specified r,

$$\frac{C\, e^{-j\beta r}}{r} = \text{constant} = C_0 \text{ (say)}$$

$$\therefore \quad E_T = C_0 I_1 e^{j\delta_1}\left(1 + \frac{I_2}{I_1}\, e^{j(\delta_2 - \delta_1)} e^{j2\pi\left(\frac{d}{\lambda}\cos\phi\right)} \right) \tag{19.66d}$$

A plot of E_T as a function of ϕ gives the radiation pattern of the array in the plane containing the axis. From this equation, the radiation pattern can be controlled by the following three parameters.

1. $\frac{d}{\lambda}$ (i.e. the ratio of the distance between the two antennae to the wavelength).
2. $(\delta_2 - \delta_1)$ (= phase difference between the two elements).
3. $\frac{I_2}{I_2}$ (i.e. the ratio of the amplitudes of the currents).

(A) Effect of Phase Difference on Radiation Pattern

Let $\frac{I_2}{I_1} = 1$ and $I_2 = I_1 = I$; then

$$E_T = C_0\, I\, e^{j\delta_1}\left(1 + e^{j(\delta_2 - \delta_1)}\, e^{j\frac{2\pi d}{\lambda}\cos\phi}\right) \tag{19.66e}$$

Since the radiation pattern is a plot of $|E|$ against the angle ϕ, we have

$$|E_T| = |C_0\, I|\left|\left(1 + e^{j(\delta_2 - \delta_1)} e^{j\frac{2\pi d}{\lambda}\cos\phi}\right)\right| \tag{19.66f}$$

Let the antenna 1 be the reference element, i.e. $\delta_1 = 0$ and the phase difference $\delta_1 - \delta_2 = \delta$. So, we get

$$|E_T| = |C_0 I|\left|\left(1 + e^{j\left(\delta + \frac{2\pi d}{\lambda}\cos\phi\right)}\right)\right|$$

$$= 2\,|C_0 I|\left|\cos\left\{\frac{1}{2}\left(\delta + \frac{2\pi}{\lambda}\cos\phi\right)\right\}\right| \tag{19.66g}$$

For maximum E_t, from the above equation,

$$\frac{1}{2}\left(\delta + \frac{2\pi}{\lambda}\cos\phi\right) = m\pi, \quad m \text{ being} = 0, 1, 2, \ldots \tag{19.66h}$$

The direction of maximum radiation is obtained from the principal value of m, i.e.

$$\left.\begin{aligned} \delta + \frac{2\pi}{\lambda}\cos\phi = 0 \\ \therefore \quad \phi_{max} = \cos^{-1}\left(\frac{-\delta\lambda}{2\pi}\right) \end{aligned}\right\} \tag{19.66j}$$

When the spacing between the elements and the wavelength of operation are specified, the direction of maximum radiation is a function of δ. So the direction of maximum radiation changes as the phase difference δ changes, i.e. when $\delta = 0$ and the two elements have the same phase,

then $\phi_{\max} = \cos^{-1}(0) = \pi/2$; or when $\delta = -\beta d = -\dfrac{2\pi d}{\lambda}$, $\phi_{\max} = \cos^{-1}(1) = 0$.

From the expression for $\phi_{\max,}$ it is seen that as δ changes from $-\beta d$ to βd, $\phi_{\max}$ changes from 0° to 180°, i.e. this is the range of variation of the direction of maximum radiation. So the inter-element phase difference is an important parameter to control the direction of maximum radiation.

(B) Inter-element Spacing (= *d*)

Let $I_1 = I_2 = I$ and $\delta_1 = \delta_2$ so that $\delta = 0$. The radiation pattern then becomes

$$|E_T| = 2|C_0 I| \left|\cos\left(\frac{\pi d}{\lambda}\cos\phi\right)\right| \tag{19.66k}$$

where $\dfrac{d}{\lambda}$ is the spacing normalized with respect to the wavelength.

$\therefore$ The condition for maximum radiation is

$$\pi\frac{d}{\lambda}\cos\phi = m\pi, \qquad m = 0, 1, 2, \ldots \text{ (integers only)}$$

$$\therefore \qquad \phi_{\max} = \cos^{-1}\left(\frac{m\lambda}{d}\right) \tag{19.66l}$$

When $\dfrac{m\lambda}{d} \le 1$, the corresponding values of $\phi_{\max}$ will be physically realizable, i.e. for all $m \le \dfrac{d}{\lambda}$, the radiation is maximum. There will be discrete number of directions of maximum radiation (= N).

$\therefore$ N = maximum permitted value of m = floor $\left(\dfrac{d}{\lambda}\right) + 1$ where floor (x) represents the highest integer $< x$.

Next, for null directions, i.e. directions along which there is no radiation, we use the same equation, allowing the cosine function to be zero and thus we get:

$$\frac{\pi d}{\lambda}\cos\phi_{\text{null}} = \left(m + \frac{1}{2}\right)\pi \qquad m = 0 \text{ and integers} \tag{19.66m}$$

So it follows that the directions of nulls and the directions of maximum radiation appear alternately [Figure 19.13(b)].

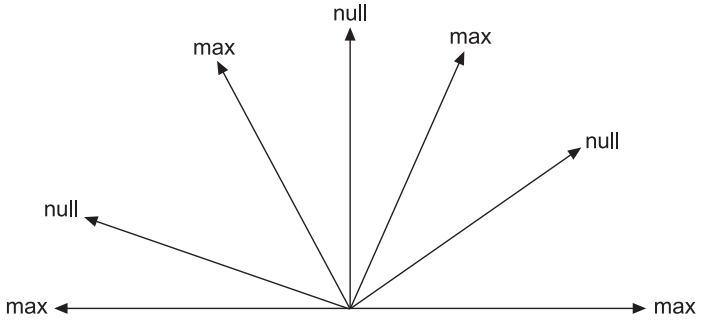


Figure 19.13(b) Maxima and nulls in a radiation pattern of a two-element array.

Thus the inter-element spacing $\frac{d}{\lambda}$ splits the radiation pattern into angular zones.

(C) Effect of Amplitude Ratio

Considering the general case, i.e. Eq. (19.66d),

$$|E| = |C_0 I_1| \left| \left\{ 1 + \frac{I_2}{I_1} e^{j\left(\delta + \frac{2\pi}{\lambda} d\cos\phi\right)} \right\} \right| \tag{19.66d}$$

let us discuss the effect of the ratio $R = \frac{I_2}{I_1}$ on the radiation pattern, allowing R to vary over the whole range 0 to ∞.

When $R = 0$, there is no excitation to the second element, and when $R \to \infty$, there is no excitation to the first element.

In both these cases, there will be no array and the radiation patterns will be those of individual antenna (i.e. isotropic in both the cases). The maximum array effect is thus realized when $I_1 = I_2$, i.e. equal excitation to both the elements. In general, the array elements are excited with equal magnitude but with different phases. There are special cases where the amplitudes of the currents have to be varied.

Thus combining the effects of these three parameters, i.e. δ, $\frac{d}{\lambda}$ and $\frac{I_1}{I_2}$ (each having different effect on the radiation pattern) a derived radiation pattern of high complexity can be achieved.

So far this analysis has been based on the assumption of isotropic antenna which is of course not physically realizable. All available antennae are non-isotropic in nature. Even the basic Hertzian dipole is non-isotropic with zero radiation along its axis and the maximum on the normal central plane. So when the elements have non-identical patterns, the radiation pattern in each plane through the array axis has to be computed, and the three-dimensional pattern would be a collection of such planar patterns.

We now consider a two-element array of non-isotropic but identical antennae, with its axis lying in the plane of the paper (say). Let the individual pattern be $f(\phi)$ in the plane of the paper. Let the two antennae be oriented identically with respect to the array axis and the excitation currents be $I_1 \angle 0$ and $I_2 \angle \delta$. The far fields at a point P, due to these two elements can be expressed as

$$\left.\begin{aligned} E_1 &= \frac{C\, I_1\, e^{-j\beta r}}{r} f(\phi) \\ E_2 &= \frac{C\, I_2\, e^{j\delta} e^{-j\beta r}}{r} f(\phi) \end{aligned}\right\} \tag{19.66n}$$

Hence the total field at the point P will be

$$E_T = \frac{C\, e^{-j\beta r} I_1}{r} f(\phi) \left(1 + \frac{I_2}{I_1} e^{j\delta} e^{j\beta d\cos\phi} \right) \tag{19.66p}$$

by analysis on similar lines as in the case of isotropic elements.

It should be noticed that Eq. (19.66p) is identical with Eq. (19.66d), except that now there is a multiplying factor $f(\phi)$.

Hence it can be stated that:

Radiation pattern of a non-isotropic array

= (Pattern of each element) × (Radiation pattern of array of isotropic elements)

↑ This is also known as Primary Pattern or Unit Pattern.

↑ This is defined by array parameters and is called Array Factor (AF).

= Element Radiation Pattern × Array Factor (= Group Pattern) (19.66q)

This is known as Pattern Multiplication.

Thus, it is possible to draw by inspection, the pattern of an array by pattern multiplication

While the 'Unit Pattern' depends on the type of the element, the 'Group Pattern' is independent of the element type so long as d, δ and orientation of the element in the array remain the same.

Note: These points which have been mentioned in general terms here, can be checked specifically for a two-element half-wave antenna discussed in Section 19.5.2 by referring to Eqs. (19.60) to (19.66).

19.5.3 Linear Arrays

When the communication has to be made from one point to another (i.e. point-to-point communication), the requisite radiation pattern is a single narrow lobe or a beam. Such a characteristic can be achieved by a multi-element linear array. An array is said to be linear, when the elements of the array are spaced equally along a straight line. A linear array is said to be uniform, when the elements are fed with currents of equal magnitude and progressive phase-shift along the line.

The pattern of such an array can be obtained by the vector addition of the field-strengths of each of the elements. For a uniform array of non-directional elements, the field-strength will be

$$E_T = E_0[1 + \exp(j\psi) + \exp(j2\psi) + \cdots + \exp(j(n-1)\psi)] \tag{19.67}$$

where

E_0 = field-strength due to the first element of the array, or due to an isolated element

$$\psi = \beta d \cos\phi + \alpha \tag{19.67a}$$

with

α = progressive phase-shift between two successive elements

d = distance between two successive elements in the array

$\beta d = 2\pi/\lambda$—path difference in radians

From Eq. (19.67),

$$\frac{E_T}{E_0} = \left|\frac{1 - \exp\{j(n\psi)\}}{1 - \exp(j\psi)}\right|$$

$$= \left| \frac{\sin\left(\frac{n\psi}{2}\right)}{\sin\left(\frac{\psi}{2}\right)} \right| \tag{19.68}$$

The maximum value of this expression is n, and this occurs when $\psi = 0$. This is the 'principal maximum' of the array. This occurs, when [from Eq. (19.67a)]:

$$\cos \phi = \frac{-\alpha}{\beta d} \tag{19.69}$$

A broadside array is one whose array factor has an absolute maximum in directions perpendicular to the axis of the array. (In our present notations, in this section, this means $\phi = \pi/2$). This means that $\alpha = 0$ for this condition. The half-power width, the full angle in which the power radiated in any direction is not less than half the maximum value, is $102°/n$, the approximation improving as n increases.

When the (currents in the) elements in an array are all in phase, but the amplitudes decrease uniformly from the central element, such an array is called a 'gabled array'. If a is the ratio of the sidelobe maximum to the main maximum of an array in which the currents are uniform, then for the corresponding gabled array, this ratio will be a^2 (as its factor is square of the uniform one). Hence the sidelobes are reduced, but the main beam is now broader, and has a half-power width of about $146°/n$ for n elements.

An end-fire array is one, which has a maximum in the direction of the axis of the array. This happens when $\phi = 0$, and hence,

$$\alpha = -\beta d \tag{19.70}$$

For this type, Eq. (19.68) is zero, when

$$\frac{n\psi}{2} = \pm k\pi, \quad k = 1, 2, 3, \ldots \tag{19.71}$$

These are the nulls of the pattern. The secondary maxima occur approximately between the zeroes or nulls, i.e.

$$\frac{n\psi}{2} = \pm (2m + 1)\frac{\pi}{2}, \quad m = 1, 2, 3, \ldots \tag{19.72}$$

Numerous graphs and details of the function (E_T/E_0) can be found in any book specifically dealing with detailed study and design of antennae (e.g. Jordan and Balmain, *Electromagnetic Waves and Radiating Systems*, 2nd Edition, Prentice-Hall of India, 2000).

Another type of linear array has elements which are called 'parasitic elements'. These elements are not connected to a generator, instead the currents in the parasitic elements are induced by the currents in the driven elements. The resultant radiation pattern is the vector sum of the patterns produced by all these elements.

A very familiar type of a linear array with parasitic elements is the Uda-Yagi array, shown in Figure 19.14. A rigorous analysis of such a system is highly complex. The distances d_1, d_2, d_3 are determined experimentally. The driven element is approximately $\lambda/2$ long, the reflector being slightly longer, and the directors shorter. The directors can be two or more.

The Uda-Yagi antenna is most often used as a receiving antenna. So far we have discussed some of the basic points of some of the fundamental types and simple and familiar types of antennae. The subject is a complex one, and for greater details and in-depth study, the interested readers are recommended texts on the subject, such as Jones, D.S., Jordan and Balmain, etc.

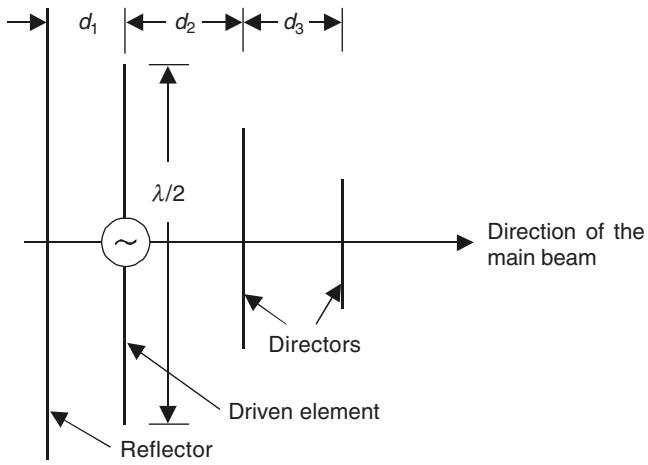


Figure 19.14 An Uda-Yagi array.

It should be noted that Eq. (19.68) is the Array Factor of the uniform linear array, i.e.

$$\text{AF} = \frac{E_\text{T}}{E_0} = \frac{\sin\dfrac{n\psi}{2}}{\sin\dfrac{\psi}{2}} \tag{19.68}$$

(or sometimes, $\text{AF} = \dfrac{E_\text{T}}{nE_0}$, where n is the number of elements in the array)

19.5.3.1 Direction of maximum radiation

The total phase difference between the fields due to any adjacent elements in the uniform linear array is

$$\psi = \beta d\cos\phi + \alpha \tag{19.67a}$$

where

$\beta d\cos\phi$ is the space phase, being a function of the direction ϕ (βd—the path difference in radians $= 2\pi/\lambda$),

α = the electrical phase – the progressive phase shift between the two successive elements, independent of the direction ϕ.

For the direction of maximum radiation, $\psi = 0$

$$\therefore \qquad \cos\phi_\text{max} = -\frac{\alpha}{\beta d}$$

or

$$\phi_\text{max} = \cos^{-1}\left(-\frac{\alpha}{\beta d}\right) = \cos^{-1}\left(\frac{-\alpha\lambda}{2\pi d}\right) \tag{19.72a}$$

This is same as Eq. (19.66j).

It is to be noticed that the direction of maximum radiation is independent of the number of elements in the uniform array. Hence as in the case of the two-element array, the direction of maximum radiation is made to vary from $\phi = 0$ to $\phi = \pi$ by allowing α to change from $-\beta d$ to $+\beta d$.

When $\phi = 0$ or $\phi = \pi$, the direction of the maximum radiation appears along the array axis, and such an array is called the 'End-fire Array'. When the direction of the maximum radiation is orthogonal to the array axis, i.e. $\phi = \pi/2$, the array is called the 'Broadside Array' [Figure 19.14(a)].

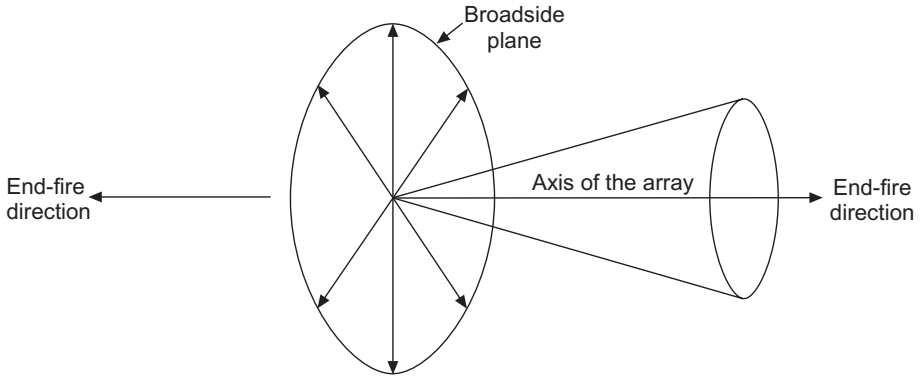


Figure 19.14(a) Broadside and end-fire directions of an antenna array.

Substituting for α (or δ of the earlier equations) in Eq. (19.67a),

$$\psi = \beta d(\cos\phi - \cos\phi_{\max}) \tag{19.67b}$$

19.5.3.2 Direction of nulls

This is the direction in which there is no radiation, i.e. the electric field is zero. Hence the required condition from Eq. (19.68) is

$$\sin\left(\frac{n\psi}{2}\right) = 0$$

or

$$\frac{n\psi}{2} = \pm m\pi, \qquad m = 1, 2, 3, \ldots$$

$$\therefore \quad \psi = \pm\frac{2m\pi}{n}, \qquad m = 1, 2, 3, \ldots \tag{19.72b}$$

$\therefore$ If ϕ_{null} denotes the direction of null

$$\psi = \beta d\cos\phi_{\text{null}} + \alpha = \pm\frac{2m\pi}{n}$$

or

$$\beta d(\cos\phi_{\text{null}} - \cos\phi_{\max}) = \pm\frac{2m\pi}{n} \tag{19.72c}$$

$$\therefore \quad \cos\phi_{\text{null}} = \cos\phi_{\max} \pm\frac{2m\pi}{\beta dn} = \cos\phi_{\max} \pm\frac{m\lambda}{dn} \tag{19.72d}$$

It should be noted that m is to be an integer $\neq 0$ because $m = 0$ corresponds to $\psi = 0$ representing maximum radiation and *not* a null. To obtain the directions of null, the integral values of m must be such that the RHS of Eq. (19.72d) must lie within the range ± 1. So there will be finite number of nulls. Also the maximum value of m with +ve and –ve signs can be different depending on the value of $\phi_{\max.}$

19.5.3.3 *Half-power beam width* (*HPBW*)

This has been defined while discussing the radiation parameters of an antenna. This is the angular separation between two directions, one on either side of the direction of maximum radiation, along which the **E** field has reduced to $\frac{1}{\sqrt{2}}$ of its maximum value. Let these two directions be ϕ_1 and ϕ_2 [Figure 19.14(b)]. Since the power density, as given by the Poynting Vector is proportional to the square of the **E** field, the power density along these two directions will be half of that along the direction $\phi_{\text{max.}}$ The angular width $(\phi_2 - \phi_1)$ is called the 'half-power beam-width' of the array.

The half-power angles ϕ_1 and ϕ_2 can be calculated from the expression for $|E|$ given in Eq. (19.68)

$$\frac{1}{n}\left|\frac{\sin\frac{n\psi}{2}}{\sin\frac{\psi}{2}}\right| = \frac{1}{\sqrt{2}} \tag{19.72e}$$

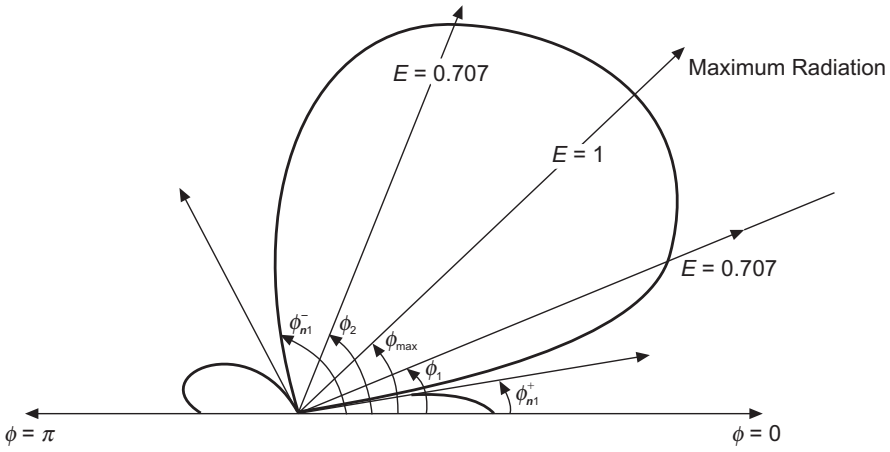


Figure 19.14(b) Half-power beam-width of a radiation pattern.

This equation has to be solved numerically to obtain the half-power angles ϕ_1 and ϕ_2, which then gives the exact value for HPBW.

Next, we have the following quick and approximate method of calculating HPBW of a large array (when an accurate answer is not required).

Note that the beam width between the first nulls (BWFN) is approximately twice the half-power beam-width (HPBW).

We assume that the first nulls are approximately symmetric with respect to the direction of maximum radiation, then

$$\left.\begin{aligned}\phi_{\text{HPBW}} = \phi_2 - \phi_1 &\simeq \frac{1}{2}(\phi_{n1}^+ - \phi_{n1}^-)\\ &\simeq (\phi_{n1}^+ - \phi_{\max})\\ &\simeq (\phi_{\max} - \phi_{n1}^-)\end{aligned}\right\} \tag{19.72f}$$

Here the implicit assumption is $\phi_{\max} \simeq \frac{1}{2}(\phi_{n1}^{+} + \phi_{n1}^{-})$

It should be appreciated that depending on the location of $\phi_{\max}$, there may be cases when ϕ_{n1}^{+} or ϕ_{n1}^{-} may not be visible in the polar plot of the radiation pattern, i.e. when $\phi_{\max}$ approaches π, the null ϕ_{n1}^{+} may not be visible, or when $\phi_{\max}$ approaches 0, ϕ_{n1}^{-} may not be visible. In such cases, the suitable expression from Eq. (19.72f) should be used.

Let us first consider the case

$$\phi_{\text{HPBW}} \simeq \phi_{n1}^{+} - \phi_{\max} \tag{19.72g}$$

then

$$\cos\phi_{n1}^{+} = \cos\phi_{\max} - \frac{\lambda}{nd} \qquad \text{from Eq. (19.72d) with } m = 1$$

or

$$\cos\phi_{n1}^{+} - \cos\phi_{\max} = -\frac{\lambda}{nd} \tag{19.72h}$$

or

$$2\sin\left(\frac{\phi_{n1}^{+} - \phi_{\max}}{2}\right)\sin\left(\frac{\phi_{n1}^{+} + \phi_{\max}}{2}\right) = \frac{\lambda}{nd}$$

or

$$2\sin\left(\frac{\phi_{\text{HPBW}}}{2}\right)\sin\left(\frac{2\phi_{\max} + \phi_{\text{PBW}}}{2}\right) = \frac{\lambda}{nd} \tag{19.72k}$$

Expanding the trigonometric function, and considering a large array for which $n >> 1$ and using the approximations $\sin x \to x$, $\cos x \to 1$ ($x << 1$) the above equation can be reduced to the following quadratic

$$\phi_{\text{HPBW}}^{2}(\cos\phi_{\max}) + (2\sin\phi_{\max})\phi_{\text{HPBW}} - \frac{2\lambda}{nd} = 0 \tag{19.72l}$$

Solving this quadratic,

$$\phi_{\text{HPBW}} = \frac{-\sin\phi_{\max} + \sqrt{\left(\sin^{2}\phi_{\max} + \frac{2\lambda}{nd}\cos\phi_{\max}\right)}}{\cos\phi_{\max}} \tag{19.72m}$$

i.e. as $\phi_{\max}$ increases from 0 to $\pi/2$, the beam width ϕ_{HPBW} decreases monotonically.

When $\phi_{\max} = 0$—the end-fire array, HPBW is maximum.

When $\phi_{\max} = \pi/2$—the broadside array, HPBW is minimum.

For these two limiting cases, from Eq. (19.72l), for the broadside array ($\phi_{\max} = \pi/2$)

$$\phi_{\text{HPBW}} = \frac{\lambda}{nd} = \frac{\lambda}{\text{length of the array}} \tag{19.72n}$$

For the end-fire array ($\phi_{\max} = 0$),

$$\phi_{\text{HPBW}} = \sqrt{\frac{2\lambda}{nd}} = \sqrt{\frac{2\lambda}{\text{length of the array}}} \tag{19.72p}$$

The length of the array is $(n - 1)d$, but when $n >> 1$, the length = nd (approx).

19.5.4 Directivity of Uniform Array

In Section 19.4.4 subsection (C), the directivity of an antenna has been defined as its power focussing capability and from the radiation pattern of an antenna, the directivity can be calculated using Eq. (19.53p).

∴ The directivity for an *N*-element uniform array,

$$D = \frac{4\pi}{\int\int_{\theta\ \phi} |\mathrm{AF}|^2\ d\Omega} \tag{19.72q}$$

where AF (= the normalized radiation pattern of the array) is given by Eq. (19.68).

For a general uniform array, this expression has to be solved numerically. For a large array, approximations are possible for the integral in the denominator. The integral can be replaced by the solid angle within the half-power beam-width of the array.

Considering the two limiting cases, i.e. broadside and end-fire arrays, and denoting their HPBW by ϕ_{BS} and ϕ_{EF} respectively.

$$\phi_{BS} = \frac{\lambda}{nd} \quad \text{and} \quad \phi_{EF} = \sqrt{\frac{2\lambda}{nd}} \tag{19.72r}$$

The solid angle for the broadside array is approximately

$$\left.\begin{aligned} \Omega_{BS} &\simeq 2\pi\phi_{BS}; \\ \text{and for the end-fire array is} \quad \Omega_{EF} &\simeq \pi\left(\frac{\phi_{EF}}{2}\right)^2 \end{aligned}\right\} \tag{19.72s}$$

The directivity of the two arrays, thus, are

$$\left.\begin{aligned} D_{BS} &= \frac{4\pi}{2\pi\phi_{BS}} = \frac{2nd}{\lambda} \\ D_{EF} &= \frac{4\pi}{\pi\left(\frac{\phi_{EF}}{2}\right)^2} = \frac{16}{\left(\frac{2\lambda}{nd}\right)^2} = \frac{8nd}{\lambda} \end{aligned}\right\} \tag{19.72t}$$

i.e. the directivity of the end-fire array is about four times that of the broadside array of same length.

On the other hand, comparing HPBWs.

$$\frac{\phi_{BS}}{\phi_{EF}} = \frac{\lambda}{nd}\sqrt{\frac{nd}{2\lambda}} = \sqrt{\frac{\lambda}{2nd}} \tag{19.72u}$$

Since $(nd) >> 1$ for large arrays, the conclusion is $\phi_{BS} >> \phi_{EF}$. This implies that the broadside array has high directivity as it has a narrower beam compared to the end-fire array.

So one has to be careful using only a planar radiation pattern. It is more important to develop the capability of visualizing the three-dimensional patterns (as three-dimensional solids).

19.6 THE RECEIVING ANTENNA

So far we have been discussing only the transmitting antennae. The task of a receiving antenna is to extract the energy from an electromagnetic wave, together with the signal carried by the wave, and deliver it to a receiver. This mechanism can be described qualitatively in simple terms. The time-varying **E** field of the wave which surrounds the antenna, induces currents and charges in the body of the conducting antenna. A certain voltage is thus developed across the antenna terminals; this voltage varying with time as per the inducing wave. This voltage is fed into the receiver which amplifies it, and then converts it into the required form which can be sound (e.g. in radio, and the corresponding audio signals of the televisions) and/or optical (e.g. optical signals in televisions, and in radars).

Though the geometrical shape of a receiving antenna is less diverse than that of a transmitting antenna, they are functionally interchangeable. But it must be remembered that the working conditions of the two types are not the same. For the transmitter, the external **E**-field wave acts only between the antenna terminals; whilst for the receiver, the external **E**-field exists all over the antenna surface. The resulting currents and the charge distributions for the two cases will be quite different.

But even then, it can be shown that the most important characteristics of the receiving antenna (which are its directive properties and its impedance) are identical to those of a transmitting antenna. In our present discussions, we shall consider the two dipoles as the two basic receiving antennae. (For further detailed information, interested readers are referred to the specialized texts mentioned earlier.)

We have seen so far that from the input side, an antenna appears as an impedance in general, the resistive part of the impedance corresponds to the power radiated by the antenna. So, for a transmitting antenna, we can say that the current (of the antenna) is the cause and the electromagnetic radiation is the effect. But for a receiving antenna the cause and the effect get reversed, i.e. the cause is the incoming electromagnetic radiation, and the effect is the current induced in the antenna surface, and the voltage induced between the antenna terminals. There are two important points to be understood regarding the receiving behaviour of the antenna: (1) Variation of the terminal voltage as a function of the direction and polarization of the incoming radiation. (2) The maximum power that can be received by the load connected to the antenna. Also of interest are the relationship between the antenna parameters in its two modes, i.e. transmitting and receiving.

The antenna is a reciprocal device (i.e. an antenna has identical radiation and circuit characteristics in transmitting and receiving modes). This follows from the reciprocity theorem which has been proved in Section 19.7.

19.6.1 The Electric and the Magnetic Dipoles (Oscillating) as Receiving Antennae

Figure 19.15 shows the electric (Hertzian) and the oscillating magnetic dipoles as the receiving antennae.

The tangential component of the incident **E** field induces currents in the wire, such that they re-radiate the energy and produce a voltage V across the load resistance R. It can be justifiably assumed that the effects of these currents on the transmitting antenna are negligible. The voltage across R can be measured by some suitable electronic voltmeter.

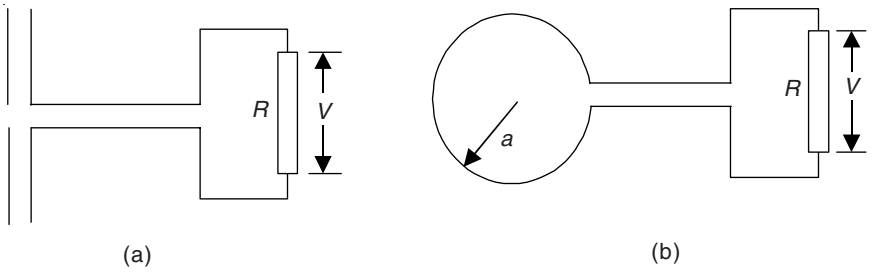


Figure 19.15 Oscillating electric and magnetic dipoles used as receiving antennae.

For the Hertzian dipole, it can be shown that,

$$V = E_t l = \mathbf{E} \cdot \mathbf{l} \; (= El \cos \alpha) \tag{19.73}$$

where

E_t = tangential component of **E** (i.e. tangential to the antenna)
l = length of the dipole antenna
V is maximal when **E** is parallel to the antenna (i.e. $\alpha = 0$).

Hertzian dipoles (or almost Hertzian dipoles) are used as receiving antennae of television receivers, portable radios, or car radios, etc.

For the magnetic dipole, the induced emf is given by

$$V = \oint_C \mathbf{E} \cdot d\mathbf{l} = \oint_C -\left(\frac{\partial \mathbf{A}}{\partial t} + \nabla V \right) \cdot d\mathbf{l}$$

$$= \iint_S -\left(\nabla \times \frac{\partial \mathbf{A}}{\partial t} + \nabla \times \nabla V \right) \cdot d\mathbf{S} \tag{19.74}$$

where S is any surface bounded by the contour C of the loop.

Now, $\nabla \times \nabla V = 0$ (a vector identity)

Also, we can interchange the order of the operators of time and space on the first term on the right. Hence, we have

$$V = \oint_C \mathbf{E} \cdot d\mathbf{l} = -\left(\frac{\partial}{\partial t} \right) \iint_S (\nabla \times \mathbf{A}) \cdot d\mathbf{S} = -\left(\frac{\partial}{\partial t} \right) \iint_S \mathbf{B} \cdot d\mathbf{S} \tag{19.75}$$

Hence, the induced emf in the loop is equal to the time-rate of change of the flux linking the loop. This is a maximum, when the normal to the loop is parallel to the local **B**. V is zero, when $\alpha = \pi/2$. Either of these two conditions can be used to determine the direction of the wave, and locate, say, a clandestine broadcasting station. Usually $\alpha = \pi/2$ condition is used, because the zero of V can be detected more accurately than a maxima. With two or more search loops, the location of a transmitting antenna can be determined with a high degree of accuracy (Figure 19.16).

It should be noted that when $R \to \infty$, then the voltage V is not necessarily the induced emf, because the circuit may be excited in the electric dipole mode. For a symmetrical loop as shown in Figure 19.15(b), when the **E** vector is in the plane of the loop, but is perpendicular to the pair of the wires leading to the loop, an extra voltage appears on R which comes from the dipole excitation and adds to the induced emf. But if the co-planar **E** is parallel to the two wires

leading to R, then the charge oscillates from one end of the wire to the other end, and V is not affected by the dipole oscillation, and so Eq. (19.75) is correct.

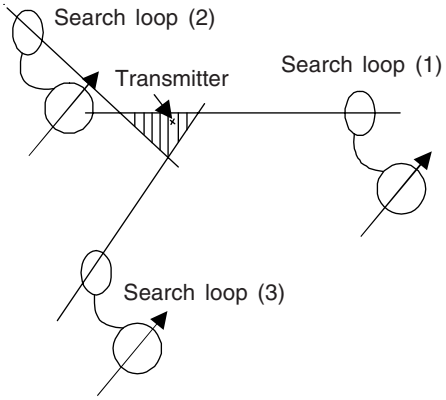


Figure 19.16 Locating a transmitter by search loop antennae.

19.6.2 Hertzian Dipole Parameters as a Receiving Antenna

(A) Radiation Pattern and Polarization

We consider a Hertzian dipole, which of course is a small length of wire with an infinitesimal gap at the centre, placed in an incoming radiation field. The dipole is located at the origin of the coordinate system, its axis being along the z-axis. A source of an electromagnetic wave is located at a very large distance R from the dipole which generates an electromagnetic wave linearly polarized in the plane of the paper.

Let this source be located at A [Figure 19.17(c)] on the y-axis (i.e. along the centre of the dipole orthogonal to its axis, i.e. $\theta = \pi/2$. So the **E** field will be polarized along the z-direction and hence parallel to the Hertz dipole.

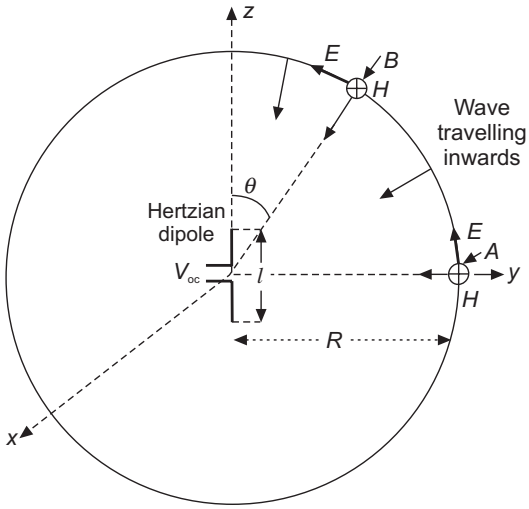


Figure 19.17(c) Hertzian dipole as a receiving antenna.

The o.c. voltage between the terminals of the dipole will be

$$V_{oc} = V_A = \mathbf{E} \cdot \mathbf{l} = El \qquad (\mathbf{E} \text{ and } \mathbf{l} \text{ are parallel}) \tag{19.75a}$$

Let the source be moved to a point B such that it is along the arc of radius R and the direction of radiation is inwards towards the Hertzian dipole.

The angle between the Elect. field $\mathbf{E}$ and the Hertzian dipole $= \frac{\pi}{2} - \theta$

$$V_{oc} = V_B = E \cdot l = El \cos\left(\frac{\pi}{2} - \theta\right) = El \sin\theta \tag{19.75b}$$

$\therefore$ Induced voltage at the receiving antenna is a function of the direction of incoming radiation θ.

When $\theta = \frac{\pi}{2}$, V_{oc} is maximum

When $\theta = 0$, $V_{oc} = 0$

Thus sin θ is the factor of dependence for V_{oc} (for a given polarization).

It should be remembered that the radiation pattern for a transmitting Hertzian dipole is $\sin\theta$. So the Hertzian dipole shows the same directional dependence for both the transmitting and the receiving mode. This is true for any radiating system, i.e. the radiation pattern for an antenna is identical for both transmitting and receiving.

We have kept the orientation of $\mathbf{E}$ unchanged in the movement of source so far. If the polarization of $\mathbf{E}$ is changed by moving the point A to the x-axis, the terminal voltage of the Hertz dipole changes and becomes zero when $\mathbf{E}$ is oriented in the x-direction. So the o.c. voltage is also a function of the polarization of the incoming radiation.

For the Hertzian dipole, V_{oc} will be maximum when $\mathbf{E}$ is along θ-direction, i.e. $\mathbf{E} = \mathbf{i}_\theta E_\theta$. So an antenna responds maximally to that polarization in the receiving mode, which it generates in the transmitting mode, i.e. the state polarization of a receiving antenna is same as that which it is capable of generating. Thus the radiation pattern and the polarization of an antenna are identical in transmitting and receiving modes.

(B) Equivalent Circuit

A transmitting antenna is equivalent to an impedance $R + jX$, where R is related to the power radiated by the antenna, and X is related to the capacitive and inductive fields surrounding the antenna.

The receiving antenna is like a voltage source with o.c. voltage V_{oc} and an internal impedance $R + jX$. The receiving antenna delivers power to the load impedance Z_L connected to its terminals. The power delivered is maximum when Z_L and Z are conjugates of each other, i.e. $Z_L = R - jX$.

Hence, the maximum power received by the antenna and delivered to the load is then

$$P_{L\max} = \frac{|V_{oc}|^2}{4R} \qquad \text{[Figure 19.17(b)]} \tag{19.75c}$$

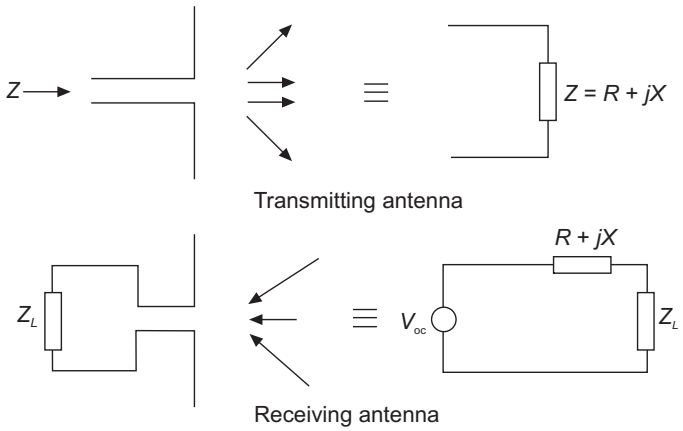


Figure 19.17(b) Equivalent circuits of the antenna.

(C) Effective Aperture (or Area)

Let the Poynting vector of the incident radiation on the receiving antenna be **S**.

$\therefore$ Power density of the incident wave = S watts/m^2

This antenna taps a power P_L from the wave.

$\therefore$ The effective aperture of the antenna $= A_e = \dfrac{P_L}{S}$ (19.75d)

This parameter is special to the receiving antenna and it gives the power capturing ability of a receiving antenna.

(D) Directive Gain

A system as shown in Figure 19.17(c) consists of a transmitting antenna 1 and a receiving antenna 2 separated by a distance r. Let the antennae 1 and 2 have internal impedances Z_1 and Z_2 respectively where antenna 2 has been connected to a conjugate load Z_2^*. From circuit considerations the two antennae would interact through a mutual impedance $Z_{\text{mu.}}$

Hence V_{oc} can be expressed as

$$V_{\text{oc}} = Z_{\text{mu}} I_1 \tag{19.75e}$$

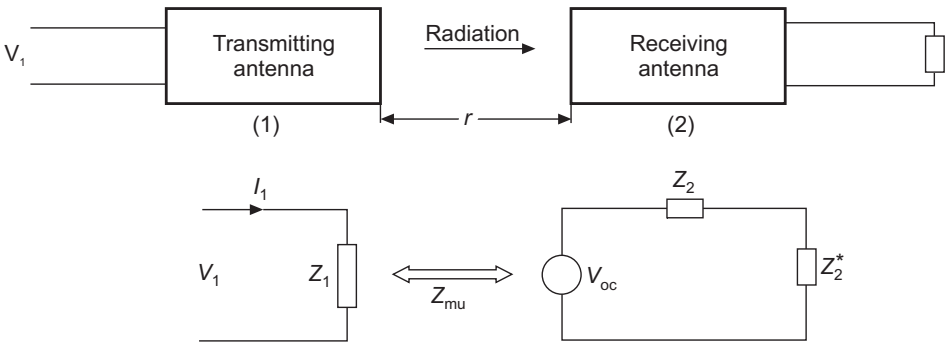


Figure 19.17(c) Transmitting and receiving antennae system.

∴ The power received by the antenna 2 and delivered to the load Z_2^* is

$$P_L = \frac{|Z_{mu}|^2 |I_1|^2}{4R_2} \tag{19.75f}$$

The power transmitted by the antenna 1 is

$$P_t = |I_1|^2 R_1 \tag{19.75g}$$

Let us assume that the gain of the transmitting antenna = G_{Dt}, and the two antennae are aligned in the direction of their maximum radiation.

∴ The power density due to antenna 1 at the location of antenna 2 is

$$S = \frac{P_t\, G_{Dt}}{4\pi r^2} = \frac{|I_1|^2 R_1 G_{Dt}}{4\pi r^2} \tag{19.75h}$$

If the effective aperture of the antenna 2 is A_{e2}, then the power received by antenna 2

$$P_r = SA_{e2} = \frac{|I_1|^2 R_1 G_{Dt}}{4\pi r^2} A_{e2} \tag{19.75j}$$

Since this is same as the power delivered to the load P_L,

then
$$P_L = P_r \rightarrow \frac{|Z_{mu}|^2 |I_1|^2}{4R_2} = \frac{|I_1|^2 R_1 G_{Dt}}{4\pi r^2} A_{e2} \tag{19.75k}$$

∴
$$|Z_{mu}|^2 = \frac{R_1 R_2 G_{Dt} A_{e2}}{\pi r^2} \tag{19.75l}$$

By the reciprocity theorem, since the two antennae are reciprocal, so

$$|Z_{mu}|^2 = \frac{R_1 R_2 G_{Dr} A_{e1}}{\pi r^2} \tag{19.75m}$$

∴ From Eqs. (19.75l) and (19.75m),

$$G_{Dt} A_{e2} = G_{Dr} A_{e1} \tag{19.75n}$$

or
$$\frac{G_{Dt}}{A_{e1}} = \frac{G_{Dr}}{A_{e2}} = K \text{ (a constant)} \tag{19.75p}$$

Since in this Section (D) no specific antenna was mentioned, Eq. (19.75p) will hold $\left(\text{i.e. the ratio } \frac{G}{A_e}\right)$ only if it is a constant, which is to say that K does not depend on any type of antenna.

(E) *K* for the Hertzian Dipole

First we calculate G for the Hertzian dipole

$$E(\theta) = \sin\theta$$

For a loss-less dipole,

$$G_D = D = \frac{4\pi}{\int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi} \sin^2\theta \sin\theta \, d\theta \, d\phi} = \frac{3}{2} \tag{19.75q}$$

Let the Hertz dipole be of length l, and be located in a radiation electric field specified as E

$$\therefore \quad \text{Power density of the wave } = S = \frac{|E|^2}{\eta} = \frac{|E|^2}{120\pi} \tag{19.75r}$$

Maximum o.c. voltage developed at the antenna terminals,

$$V_{oc} = El$$

Note: Radiation resistance of the Hertzian dipole = $R_{rad} = \dfrac{80\pi^2 l^2}{\lambda^2}$

$\therefore$ Maximum power delivered (to the matched load)

$$P_L = \frac{V_{oc}^2}{4R_{rad}} = \frac{|E|^2 l^2}{4 \times 80\pi^2}\left(\frac{\lambda^2}{l^2}\right) \tag{19.75s}$$

$\therefore$ The effective aperture for the Hertez dipole

$$A_e = \frac{P_L}{S} = \frac{3\lambda^2}{8\pi} \tag{19.75t}$$

$\therefore$ Effective aperture of Hertzian dipole is not dependent on its length and is a function of the operating wavelength only.

$$\therefore \quad K = \frac{G_D}{A_e} = \frac{4\pi}{\lambda^2} \tag{19.75u}$$

For a general antenna,

$$G_D = \frac{4\pi A_e}{\lambda^2} \tag{19.75v}$$

It should be noted that the 'Directive Gain' is a parameter of a transmitting antenna, whereas the 'Effective Aperture' is a parameter of a receiving antenna. So for an antenna which can operate in both modes (by the theorem of reciprocity), these two parameters, G_D and A_e, are directly proportional, i.e. if an antenna has high gain while transmitting, its effective aperture will also be high while in receiving mode.

Referring back to 2-antennae system of (D), from Eq. (19.75h), we have

$$P_t = \frac{4\pi r^2 S}{G_{Dt}}$$

From Eq. (19.75d),

$$P_L = A_{e2}\, S = P_r$$

which combined with Eq. (19.75v), gives

$$P_r = S \cdot \frac{\lambda^2}{4\pi} G_{Dr}$$

$$\therefore \quad \frac{P_r}{P_t} = \frac{S\lambda^2 G_{Dr}/(4\pi)}{4\pi r^2 S} = G_{Dt}\, G_{Dr} \left(\frac{\lambda}{4\pi r}\right)^2$$

$$\therefore \quad P_r = G_{Dt} G_{Dr} \left(\frac{\lambda}{4\pi r}\right)^2 P_t \tag{19.75w}$$

This is known as 'Friis Transmission Formula' relating the power received by one antenna to the power transmitted by another provided that the two antennae are separated by a distance $r > 2d^2/\lambda$ where d is the largest dimension of either antenna. So care must be taken to see that the two antennae are in the far field of each other.

19.7 THE RECIPROCITY THEOREM

It is known, in the theory of electrostatics, that certain relations exist between two possible potential and charge distributions on a given system of conductors (see Appendix 3). Such relations are commonly known as the 'reciprocity theorems', and similar theorems exist for the harmonic electromagnetic field. Each component of the field vector satisfies the harmonic wave equation,

$$\nabla^2\phi + k^2\phi = 0 \tag{19.76}$$

However, we start by discussing a more general equation, which can be written as

$$\text{Div } (a \text{ grad } f) + k^2 f = 0 \tag{19.77}$$

where a and k are known functions. When $a = 1$, this equation reduces to Eq. (19.76). Let T be the volume enclosed by a closed surface S, and let f_1 and f_2 be any two solutions of the Eq. (19.77) which satisfy the same boundary conditions at any boundaries within T.

$\therefore$ It follows from the divergence theorem, that

$$\oiint_S a \left\{ f_1 \left(\frac{\partial f_2}{\partial n}\right) - f_2 \left(\frac{\partial f_1}{\partial n}\right) \right\} dS = 0 \tag{19.78}$$

provided that the integrand is continuous across all boundaries within T.

The boundary conditions of this type are:

1. f and $(a\ \partial f/\partial n)$ are continuous, or
2. (af) and $(\partial f/\partial n)$ are continuous.

The more general boundary condition containing both these conditions is:

$a\{(\partial f/\partial n) + \alpha f\}$ and $\{\beta\,(\partial f/\partial n) + f\}$ are continuous, where α and β are continuous, and such that $(\alpha\,\beta) \neq 1$.

The continuity of the integrand with these conditions can be seen by writing it in the form:

$$\left(\frac{a}{1-\alpha\beta}\right)\left[\left\{f_1+\beta\left(\frac{\partial f_1}{\partial n}\right)\right\}\left\{\left(\frac{\partial f_2}{\partial n}\right)+\alpha f_2\right\}-\left\{f_2+\beta\left(\frac{\partial f_2}{\partial n}\right)\right\}\left\{\left(\frac{\partial f_1}{\partial n}\right)+\alpha f_1\right\}\right]$$

This theorem holds also if the boundary condition is $f = 0$ or $(\partial f/\partial n) = 0$ or more generally $(\partial f/\partial n) + \sigma f = 0$. Equation (19.78), which is valid under the appropriate boundary conditions, is one form of the reciprocity theorem. This form does not include the sources in the region. A different form of the theorem is for when the sources are present in the region under consideration, i.e. inside T. In this case, let $f_1(R)$ and $f_2(R)$ satisfy the equations,

$$\text{div } (a \text{ grad } f_1) + k^2 f_1 = -\delta(R - R_1) \tag{19.79a}$$

$$\text{div } (a \text{ grad } f_2) + k^2 f_2 = -\delta(R - R_2) \tag{19.79b}$$

where $R_1 \neq R_2$, and R_1 and R_2 are points of T. Then

$$\oiint_S a\left[f_1\left(\frac{\partial f_2}{\partial n}\right) - f_2\left(\frac{\partial f_1}{\partial n}\right)\right] dS = f_2(R_1) - f_1(R_2) \tag{19.80}$$

provided f_1 and f_2 satisfy the boundary conditions of the type mentioned earlier. If, in addition,

$$\frac{\partial f_1}{\partial n} + \sigma f_1 = 0, \quad \frac{\partial f_2}{\partial n} + \sigma f_2 = 0, \quad \text{on } S,$$

then

$$f_1(R_2) = f_2(R_1) \tag{19.81}$$

Under these conditions, the reciprocity theorem states that:

The field produced at the second source by the first source is the same as that produced at the first source by the second.

In case Eqs. (19.77) and (19.79) are not satisfied inside S, but valid outside, then the reciprocity theorems (19.78), (19.80), and (19.81) are still valid provided f_1 and f_2 satisfy suitable radiation conditions at infinity.

The corresponding theorems for the electromagnetic fields are obtained in a similar manner.

This theorem when applied to electromagnetic waves states that 'the current in a detector divided by the voltage at the source remains constant when the source and the detector are interchanged, so long as the frequency and all the impedances remain unchanged'.

This theorem is very widely used for investigating circuits as well as antennae. Here we shall prove the generalized form by using the Maxwell's equations.

To start with, we have a pair of loop antennae, one of which is being used as a transmitter, and the other is used as a receiver, as shown in Figure 19.18 shown below.

The conductors and the medium of propagation are assumed to be isotropic. The source connected to the left-hand antenna supplies a voltage V, and the detector is connected to the right-hand antenna and measures a current I.

By the reciprocity theorem, (I/V) remains unchanged, if the source and the detector are interchanged, as shown in Figure 19.19. We now consider the field a (i.e. $\mathbf{E}_a$, $\mathbf{H}_a$) that is obtained when the antenna a is the transmitter on being connected to the oscillator (the top part of the Figure 19.19), and then the different field b (i.e. $\mathbf{E}_b$, $\mathbf{H}_b$) when the antenna b is made the transmitter by connecting it to the oscillator as shown in the lower part of Figure 19.19. We assume the frequency and the impedances to remain unchanged for these two arrangements.

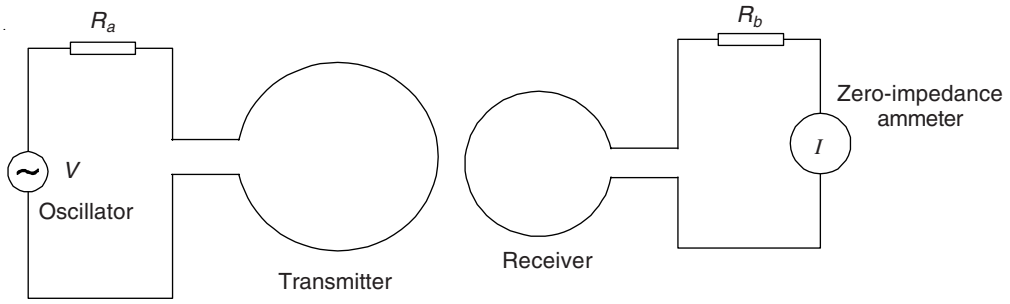


Figure 19.18 A pair of loop antennae. The one on the left (transmitter) is excited by an oscillator, supplying a voltage V. The other (receiver) is connected to a load resistance R_b through a zero-impedance ammeter, measuring the current I.

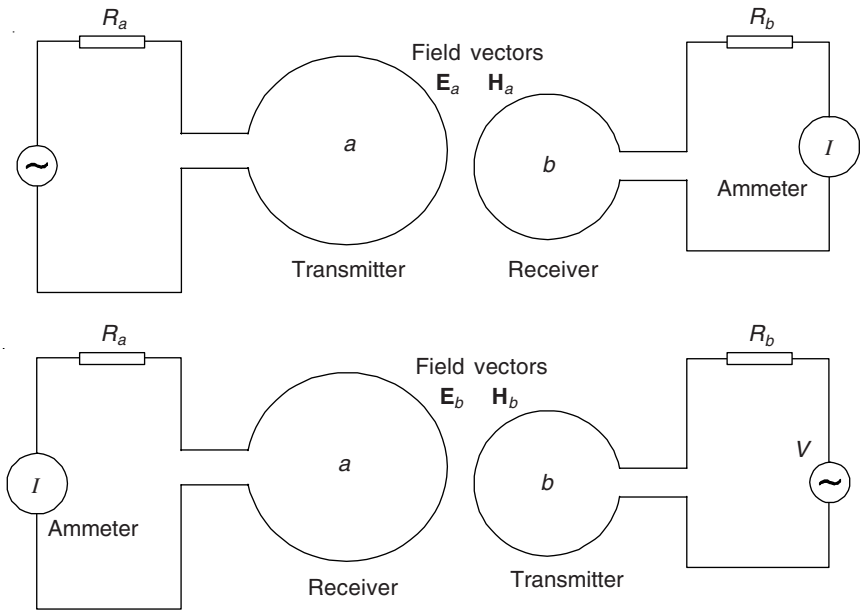


Figure 19.19 A pair of loop antennae with the source in a at the top, and then in b (bottom). The frequency and the impedances remain unchanged.

Then for any point in the space which includes the antennae as well as the source, the following vector identity holds:

$$\nabla \cdot (\mathbf{E}_a \times \mathbf{H}_b - \mathbf{E}_b \times \mathbf{H}_a) = \mathbf{H}_b \cdot (\nabla \times \mathbf{E}_a) - \mathbf{E}_a \cdot (\nabla \times \mathbf{H}_b) - \mathbf{H}_a \cdot (\nabla \times \mathbf{E}_b) - \mathbf{E}_b \cdot (\nabla \times \mathbf{H}_a) \quad (19.82)$$

Using the Maxwell's equations,

$$\begin{aligned} \nabla \cdot (\mathbf{E}_a \times \mathbf{H}_b - \mathbf{E}_b \times \mathbf{H}_a) = & -\mathbf{H}_b \cdot \left(\frac{\partial \mathbf{B}_a}{\partial t} \right) - \mathbf{E}_a \cdot \left(\mathbf{J}_{fb} + \frac{\partial \mathbf{D}_b}{\partial t} \right) \\ & + \mathbf{H}_a \cdot \left(\frac{\partial \mathbf{B}_b}{\partial t} \right) + \mathbf{E}_b \cdot \left(\mathbf{J}_{fa} + \frac{\partial \mathbf{D}_a}{\partial t} \right) \end{aligned} \quad (19.83)$$

Since we are considering the time-harmonic fields, $\frac{\partial}{\partial t} \equiv j\omega$,

and on simplifying, this equation reduces to:

$$\nabla \cdot (\mathbf{E}_a \times \mathbf{H}_b - \mathbf{E}_b \times \mathbf{H}_a) = \mathbf{E}_b \cdot \mathbf{J}_{fa} - \mathbf{E}_a \cdot \mathbf{J}_{fb} \tag{19.84}$$

For points outside the source, $\mathbf{J}_f = \sigma\mathbf{E}$ (assuming that the Ohm's law holds). But within the source there is another electric field intensity $\mathbf{E}_s$, and hence, in general:

$$\mathbf{J}_{fa} = \sigma(\mathbf{E}_a + \mathbf{E}_{sa}) \tag{19.85a}$$

$$\mathbf{J}_{fb} = \sigma(\mathbf{E}_b + \mathbf{E}_{sb}) \tag{19.85b}$$

The quantities $\mathbf{E}_{sa}$ and $\mathbf{E}_{sb}$ are the applied field intensities within the source when it is in the loop a, and when it is in the loop b, respectively.

$\mathbf{J}_{fa}$ = the current density at any point in space, when the source is in a
$\mathbf{J}_{fb}$ = the current density at any point in space, when the source is in b.

$\therefore$ Substituting these in Eq. (19.84), we obtain

$$\nabla \cdot (\mathbf{E}_a \times \mathbf{H}_b - \mathbf{E}_b \times \mathbf{H}_a) = \mathbf{E}_{sa} \cdot \mathbf{J}_{fb} - \mathbf{E}_{sb} \cdot \mathbf{J}_{fa} \tag{19.86}$$

In general, the right-hand side of the equation is not equal to zero.

This above relationship holds for any pair of electromagnetic fields at any points in space, including the region inside the sources. Hence, integrating over the whole space,

$$\iiint_{\infty} \nabla \cdot (\mathbf{E}_a \times \mathbf{H}_b - \mathbf{E}_b \times \mathbf{H}_a)\, dv = \iiint_{\infty} (\mathbf{E}_{sa} \cdot \mathbf{J}_{fb} - \mathbf{E}_{sb} \cdot \mathbf{J}_{fa})\, dv \tag{19.87}$$

By applying the divergence theorem,

$$\oiint_{\infty} (\mathbf{E}_a \times \mathbf{H}_b - \mathbf{E}_b \times \mathbf{H}_a)\, d\mathbf{S} = \iiint_{\infty} (\mathbf{E}_{sa} \cdot \mathbf{J}_{fb} - \mathbf{E}_{sb} \cdot \mathbf{J}_{fa})\, dv \tag{19.88}$$

If the sources are now constrained to a finite volume, the surface of integration on the left-hand side is infinitely away from them, and then there is a plane wave with $\mathbf{E}$ and $\mathbf{H}$ orthogonal and transverse.

$\therefore$ $$\mathbf{E} \times \mathbf{H} = \mathbf{r}_1 EH \tag{19.89}$$

where $\mathbf{r}_1$ is the unit radial vector.

$\therefore$ $$\mathbf{H}_a = \left(\frac{1}{Z_0}\right)\mathbf{r}_1 \times \mathbf{E}_a \quad \text{and} \quad \mathbf{H}_b = \left(\frac{1}{Z_0}\right)\mathbf{r}_1 \times \mathbf{E}_b \tag{19.90}$$

$\therefore$ The integrand on the left-hand side of Eq. (19.88) is

$$(\mathbf{E}_a \times \mathbf{H}_b - \mathbf{E}_b \times \mathbf{H}_a) = \left(\frac{1}{Z_0}\right)[\mathbf{E}_a \times (\mathbf{r}_1 \times \mathbf{E}_b) - \mathbf{E}_b \times (\mathbf{r}_1 \times \mathbf{E}_a)] = 0 \tag{19.91}$$

at points infinitely remote from the sources.

$\therefore$ The integral on the right-hand side of Eq. (19.88) must also be zero, i.e.

$$\iiint_{\infty} (\mathbf{E}_{sa} \cdot \mathbf{J}_{fb} - \mathbf{E}_{sb} \cdot \mathbf{J}_{fa})\, dv = 0 \tag{19.92}$$

The integration is extended to all space, but it can be restricted to the sources, since $\mathbf{E}_{sa}$ and $\mathbf{E}_{sb}$ are zero elsewhere.

$$\therefore \quad \iiint_{\infty} (\mathbf{E}_{sa} \cdot \mathbf{J}_{fb})\, dv = \iiint_{\infty} (\mathbf{E}_{sb} \cdot \mathbf{J}_{fa})\, dv \tag{19.93}$$

where the integrals are evaluated over the regions where $\mathbf{E}_{sa}$ and $\mathbf{E}_{sb}$ are non-zeroes.

Physically what this equation means, can be explained in terms of the pair of the loop antennae a and b, i.e.

$$\iiint_{a} (\mathbf{E}_{sa} \cdot \mathbf{J}_{fb}) dv = \iiint (\mathbf{E}_{sa} \cdot d\mathbf{l})(\mathbf{J}_{fb} \cdot d\mathbf{S}) = V_{sa} I_{b \text{ in } a}$$

where V_{sa} is the voltage supplied by the source in a, and $I_{b \text{ in } a}$ is the current in the same loop a when b is energized.

Similarly,

$$\iiint_{b} (\mathbf{E}_{sb} \cdot \mathbf{J}_{fa})\, dv = \iiint (\mathbf{E}_{sb} \cdot d\mathbf{l})\, (\mathbf{J}_{fa} \cdot d\mathbf{S}) = V_{sb} I_{a \text{ in } b}$$

$$\therefore \quad V_{sa} I_{b \text{ in } a} = V_{sb} I_{a \text{ in } b}$$

$$\therefore \quad \frac{I_{a \text{ in } b}}{V_{sa}} = \frac{I_{b \text{ in } a}}{V_{sb}}$$

i.e. the physical interpretation is:

> *The current induced in b when a is energized, divided by the voltage applied on a is equal to the current induced in a when b is energized, divided by the applied voltage on b (keeping the frequency and the impedances unchanged).*

This is the reciprocity theorem. It is valid for any pair of antennae. It should also be noted that this theorem is concerned only with the ratio (I/V), and does not say anything about the power expended by the source (which usually changes when the source is moved from one position to the other).

PROBLEMS

19.1 The field near to a Hertzian dipole of length l has the following principal components in spherical polar coordinates:

$$E_r = \frac{ql \cos\theta}{2\pi\varepsilon_0 r^3}; \qquad E_\theta = \frac{ql \sin\theta}{4\pi\varepsilon_0 r^3}; \qquad B_\phi = \frac{\mu_0 il \sin\theta}{4\pi r^2}$$

If i is oscillating and equal to $I\sqrt{2} \cos \omega t$, prove that the predominant energy flow in this region is likewise oscillatory, being such that a quantity of energy given by

$$W = \frac{i^2 l^2}{6\pi\varepsilon_0 \omega^2 r^3}$$

flows out and back from a sphere of radius r, twice in each cycle of the dipole current.

19.2 Show that the phase velocity of **H** field of an oscillating dipole is

$$v_\phi = c\left(1 + \frac{c^2}{\omega^2 r^2}\right)$$

Show also that the phase velocities of r- and θ-components of **E** field are:

$$v_r = c\left(1 + \frac{c^2}{\omega^2 r^2}\right), \quad \text{and} \quad v_\theta = c\left[\frac{(\omega r/c)^4 - (\omega r/c)^2 + 1}{(\omega r/c)^4 - 2(\omega r/c)^2}\right] \text{ respectively.}$$

(The charges of the dipole are $\pm Qe^{j\omega t}$.)

19.3 The symmetry of Maxwell's equations in free space implies that any system of travelling waves defined by the field vectors **E**, **B**, has a dual in which $\mathbf{E}' = -c\mathbf{B}$, $\mathbf{B}' = \mathbf{E}/c$. What source would produce a field which is the dual of that set up by a Hertzian dipole?

19.4 Discuss and draw the image of a horizontal dipole antenna above a perfectly conducting plane, and show that the current in the image and the current in the antenna flow in *opposite* directions.

Discuss and draw again the image of a vertical dipole (antenna) above a perfectly conducting plane, and show that in this case, the current in the image as well as in the dipole flow in the *same* direction.

19.5 Considering the far fields of the electric dipole and the magnetic dipole, show that they are duals of each other.

19.6 The field of a magnetic dipole is such that $V = 0$ and $\mathbf{A} \neq 0$. Is it possible to have a radiation field which has $V \neq 0$, $\mathbf{A} = 0$?

19.7 A sealed plastic box contains a transmitting antenna radiating electromagnetic waves. How would you identify whether it is a magnetic or electric dipole?

20 Electromagnetism and Special Relativity

20.1 INTRODUCTION

So far, in our study of electromagnetism, whenever we have distinguished the 'electrostatic' field of stationary charges from the 'electromagnetic' field of moving charges, we have made a tacit assumption that the earth is fixed, and 'stationary' means at rest relative to the earth's surface. But the earth is not fixed in any absolute sense, and the dominant factor is whether there exists rest or motion with respect to a particular observer. Electromagnetic observations, made by two observers in relative motion, will differ. However before we apply the relativistic concepts to the electromagnetic phenomena, we shall have a look at the concepts of relativity in the context of classical mechanics.

Furthermore, if we have a look at the historical sequence of the scientific developments, we find that Maxwell's electromagnetic theory (1855–1865) was based on the experimental laws of Coulomb, Ampere, and Faraday. The special theory of relativity was enunciated in 1904–1905. It was L. Page, who in 1912 said that 'if the principle of special relativity had been enunciated before the date of Oersted's discovery of the magnetic effects of electric currents, then the fundamental relations of Maxwellian electrodynamics could have been predicted on theoretical grounds as a direct consequence of the fundamental laws of electrostatics, extended so as to apply to charges in relative motion, as well as to charges at rest'.

We shall try to show in this chapter that starting from electrostatics (i.e. Coulomb's law), and applying the transformations of special relativity, we can derive Ampere's, Faraday's laws, and the generalization into Maxwell's equations.

This will give us a better insight and deeper understanding of all aspects of the electromagnetic theory, and thereby lay the groundwork for more advanced study of electrodynamics in moving media. It will also help us in finding correct solutions for indeterminate or ambiguous problems caused by the flux-cutting rule. Finally, various isolated laws will appear as different facets of a single comprehensive theory. However, before we apply the concepts of relativity to electromagnetic phenomena, we shall have a look at some of the simpler phenomena in classical mechanics so as to be clear about the 'relative velocity'.

20.2 GALILEIAN RELATIVITY (GALILEO GALILIE)

This is the simplest case in classical mechanics, so much so that it is applied without any conscious thought process.

Let us consider a boat, on a river moving downstream with a velocity v relative to the water of the river, and that the river water has a velocity u (in the same direction).

Hence the velocity of the boat relative to a stationary observer on the bank of the river = v + u.

And (to an observer sitting on the boat), the observer on the bank appears to move with a velocity equal to = −(v + u).

∴ The position of the boat can be specified with respect to:

1. a fixed point on the bank of the river, or
2. a point on the river—moving with the water, i.e. a drifting bouy.

So now we choose a coordinate system S fixed to a point which is fixed on the bank, and another coordinate system S' whose origin is fixed on the drifting bouy (Figure 20.1).

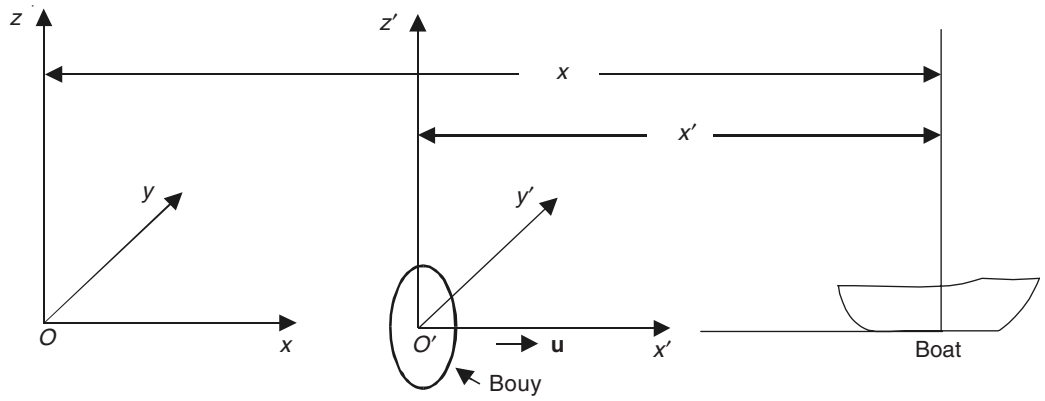


Figure 20.1 Coordinate systems S and S'.

Hence,

x = the position of the boat with respect to the coordinate system S; and
x' = the position of the boat with respect to the coordinate system S'; then we can write:

$$x' = x - \mathrm{u}t \tag{20.1}$$

This equation expresses the position of the boat relative to the drifting bouy in terms of the position and the time measured by an observer fixed on the bank.

Or, we could equally well have written this relationship in the form

$$x = x' + \mathrm{u}t', \quad (t = t') \tag{20.2}$$

where the position of the boat relative to the bank has now been expressed in terms of the position and the time measured by an observer sitting on the bouy and drifting along with it.

This transformation from one coordinate system to another as given by Eq. (20.1) or (20.2) is known as the Galilean transformation, named after Galileo Galilei. In classical Newtonian mechanics, the basic hypothesis is that the laws of motion hold equally in all rigid coordinate

frames moving with uniform rectilinear velocity with respect to each other. The equations of motion remain unchanged in the two systems by using the two transformation relations mentioned above. This invariance of the laws of mechanics under a Galilean transformation is called the 'Galilean relativity'. There are certain assumptions implicit in this transformation. They are that the mass and the dimensions remain same at rest as under motion. Also, it is important to note that the variation of time is at the same rate for both moving and fixed observers. As we shall see later, that these assumptions are valid only when the velocities under consideration are much smaller when compared with the velocity of light (= c), i.e. v << c. These laws become different when we have much higher velocities, i.e. the velocities comparable to that of the light velocity.

20.3 ELECTROMAGNETIC PHENOMENA AS VIEWED BY DIFFERENT OBSERVERS

We have two inertial frames of reference F and F' where the frame F' moves with a constant velocity u in the x-direction with respect to the frame F.
(*Note:* We define an 'inertial frame of reference' as follows: 'If in a reference frame, a particle under the influence of no forces travels in a straight line with constant speed, then that reference frame is called an inertial frame of reference'. The laboratory system is generally a very satisfactory approximation to an inertial reference frame for the description of electromagnetic phenomena.)

The two frames of reference F and F' have a common x-axis as shown in Figure 20.2 below.

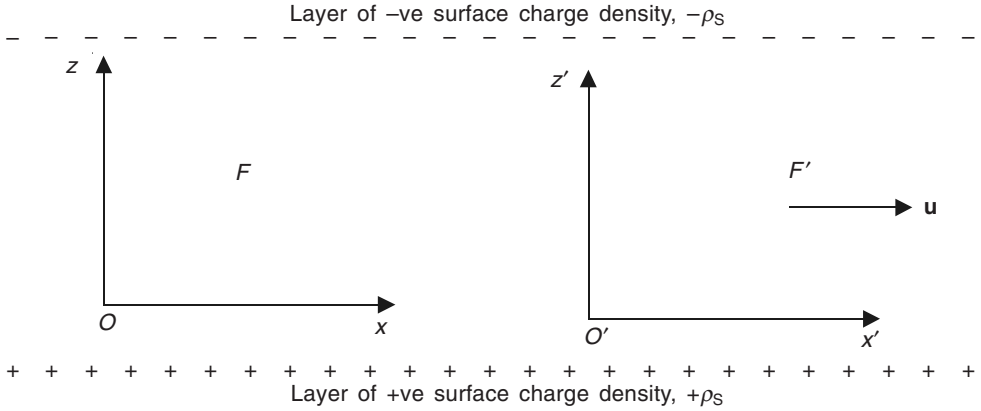


Figure 20.2 Parallel layers of electric charge.

There exist two layers of electric charges parallel to the x-y plane, y-axis being normal (into) the plane of the paper. We consider an observer making measurements at O, a point between the two sheets of charges (as shown above) of densities $\pm \rho_S$ per unit area. Since the electric field is stationary in this frame, the observer will detect an electric field having one component given by

$$E_z = \frac{\rho_S}{\varepsilon_0} \tag{20.3}$$

and no magnetic field.

In the other frame of reference F' with the axes $O'\,x'\,y'\,z'$, moving relatively to F with the velocity u in the direction Ox, an observer (sitting at O') will also detect the electric field E_z, but for him the charges are gliding backwards with the velocity u, and so the upper charge-sheet appears as a current-sheet of (ρ_Su) amps/metre directed in the direction $O'x'$ whilst the lower charge-sheet appears as a similar current-sheet in the reverse direction. The diagram (Figure 20.3)

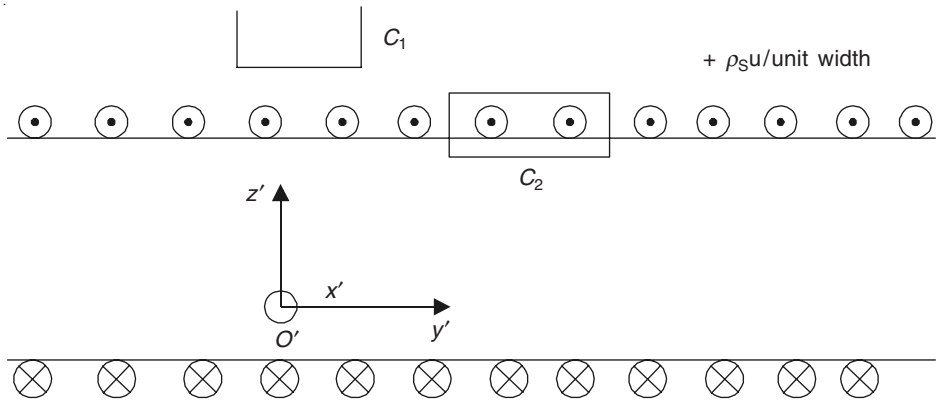


Figure 20.3 End-view of the charge-sheets as seen by an observer in the moving frame of reference F'.

shows the end-view of these currents. At infinity they seem coincident, so produce no magnetic field, and hence a magnetic circuit such as C_1 shows that there exists no magnetic field anywhere outside the currents, while the circuit C_2 shows that between them (the current-sheets—as they appear to the observer in the moving frame of reference), there exists a magnetic field in the direction $O'y'$, which is given by

$$H'_y = \mathrm{u}\rho_S = -J'_{Sx} \tag{20.4}$$

or

$$B'_y = \mu_0\rho_S\mathrm{u} = \mu_0\varepsilon_0\mathrm{u}E_z = \left(\frac{\mathrm{u}}{\mathrm{c}^2}\right)E_z \tag{20.5}$$

Thus the observer O' in the frame F' observes this magnetic field as well as the electric field which the observer O in the frame F discerns, i.e.

$$E'_z = E_z \tag{20.6}$$

Let the charge-sheets now be replaced by iron poles with respect to which the frame F is at rest. Let the magnetic field between these poles be observed from F as B_z (Figure 20.4).

If the frame F' carried a conductor along the axis $O'y'$, then an emf ($\mathrm{u}B_z$) would be found to be induced in it in the direction $O'y'$. This emf is a sign that if the conductor were absent, there would be an electric field given by

$$E'_y = -\mathrm{u}B_z \tag{20.7}$$

The observer in the frame F' will observe this field in addition to the magnetic field B_z.

The magnetic flux density B'_y given by the Eq. (20.5), introduced into the observations on the moving frame F' by its relative motion to the frame F, is got by multiplying (F)'s electric field E_z by (u/c^2) and swinging the vector through negative 90° about the direction of u. If (F)'s field had contained a component E_y, the same rule would give

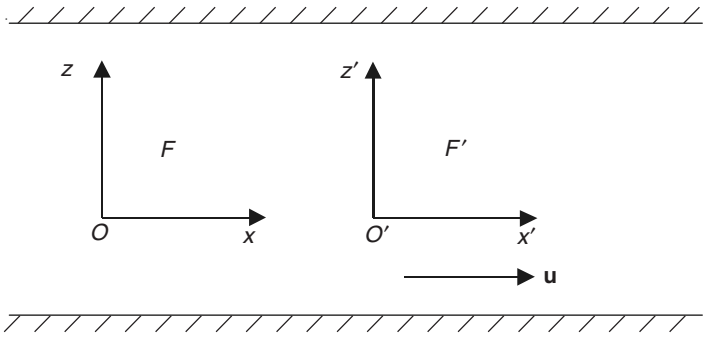


Figure 20.4 Two parallel magnetic poles.

$$B'_z = -\left(\frac{\mathrm{u}}{\mathrm{c}^2}\right)E_y \tag{20.8}$$

Similarly the electric force E'_y given by Eq. (20.7) is obtained by multiplying the flux density B_z observed from the frame F by u and swinging the vector through positive 90° about the direction of u. So by similar arguments [as we had for Eq. (20.8)], if the frame (F)'s field had contained a component B_y, then the same rule (mentioned before in this paragraph) would give

$$E'_z = \mathrm{u}B_y \tag{20.9}$$

But the components in the x-direction (i.e. the direction of motion) would be observed equally from both the frames. Hence

$$E'_x = E_x \quad \text{and} \quad B'_x = B_x \tag{20.10}$$

Thus building up a general field by superposition of perpendicular components. we find that the general relation between the vectors **E**, **B** in the frame F and **E′**, **B′** in the frame F' which has a relative motion u with respect to F in the direction Ox, is contained in these equations:

$$E'_x = E_x, \qquad E'_y = E_y - \mathrm{u}B_z, \qquad E'_z = E_z + \mathrm{u}B_y \tag{20.11a}$$

$$B'_x = B_x, \qquad B'_y = B_y + \left(\frac{\mathrm{u}}{\mathrm{c}^2}\right)E_z, \qquad B'_z = B_z - \left(\frac{\mathrm{u}}{\mathrm{c}^2}\right)E_y \tag{20.11b}$$

These equations have been derived with the assumption of uniform fields, but with fields which vary from place to place, they are still true at the instant when the frames F and F' coincide. Equations (20.11) can be written in compact vector form as:

$$\mathbf{E}' = \mathbf{E} + (\mathbf{u} \times \mathbf{B}) \tag{20.12a}$$

$$\mathbf{B}' = \mathbf{B} - \left(\frac{1}{\mathrm{c}^2}\right)(\mathbf{u} \times \mathbf{E}) \tag{20.12b}$$

As a first application, we may get the formula for the force on a moving charge. Let the charge q move with a velocity v = u in the frame F in which exists both an electric field **E** and a magnetic field **B**. Let q move with the origin of F', it is then static in F', and so the force on it is given by

$$\mathbf{P}' = q\mathbf{E}' \tag{20.13}$$

Transforming to the frame F, and assuming:

1. q is same for both the frames (true);
2. $\mathbf{P} = \mathbf{P}'$ (true for the order of the accuracy here considered),

we get [from Eq. (20.12)]

$$\mathbf{P} = q[\mathbf{E} + (\mathbf{u} \times \mathbf{B})] \tag{20.14}$$

This is the important *Lorentz Force Formula*, which we had used to define the magnetic flux density vector **B**.

20.4 TRANSFORMATION OF ELECTRIC AND MAGNETIC FIELDS

Equations (20.11) were derived from the Galilean relativity which is applicable to the Newtonian mechanics.

∴ Similar to these principles, the physical results obtained should be independent, whether the frame F or F' is regarded as fixed, i.e. it is valid to consider F' as fixed and F moving in the x' direction with a velocity $(-\mathrm{u})$.

∴ The unprimed fields should be obtained from the primed fields by a set of relations corresponding exactly to Eqs. (20.11) with the primed and the unprimed quantities interchanged, and u replaced by $(-\mathrm{u})$.

∴ Equations (20.11) should transform to

$$E_x = E'_x, \qquad E_y = E'_y + \mathrm{u}B'_z, \qquad E_z = E'_z - \mathrm{u}B'_y \tag{20.15a}$$

$$B_x = B'_x, \qquad B_y = B'_y - \left(\frac{\mathrm{u}}{\mathrm{c}^2}\right)E'_z, \qquad B_z = B'_z + \left(\frac{\mathrm{u}}{\mathrm{c}^2}\right)E'_y \tag{20.15b}$$

But, if Eqs. (20.11) are solved to obtain the unprimed field quantities, then the following set of relations is obtained by the steps which are obtained from Eqs. (20.12), i.e. Eqs. (20.12) can be re-written as

$$\begin{aligned}\mathbf{E} &= \mathbf{E}' - (\mathbf{u} \times \mathbf{B}) \\ &= \mathbf{E}' - (\mathbf{u} \times \mathbf{B}') - \left(\frac{1}{\mathrm{c}^2}\right)[\mathbf{u} \times (\mathbf{u} \times \mathbf{E})]\end{aligned}$$

If **u** has only x-component, the components of $(\mathbf{u} \times \mathbf{E})$ are 0, $-\mathrm{u}E_z$, $\mathrm{u}E_y$; and of $[\mathbf{u} \times (\mathbf{u} \times \mathbf{E})]$ are 0, $-\mathrm{u}^2E_y$, $-\mathrm{u}^2E_z$.

Thus the equations for **E** and **B** in terms of **E′** and **B′** are

$$E_x = E'_x \tag{20.16a}$$

$$E_y = \frac{E'_y + \mathrm{u}B'_z}{1 - \dfrac{\mathrm{u}^2}{\mathrm{c}^2}} \tag{20.16b}$$

$$E_z = \frac{E'_z - \mathrm{u}B'_y}{1 - \dfrac{\mathrm{u}^2}{\mathrm{c}^2}} \tag{20.16c}$$

and

$$B_x = B'_x \tag{20.17a}$$

$$B_y = \frac{B'_y - \left(\frac{u}{c^2}\right)E'_z}{1 - \frac{u^2}{c^2}} \tag{20.17b}$$

$$B_z = \frac{B'_z + \left(\frac{u}{c^2}\right)E'_y}{1 - \frac{u^2}{c^2}} \tag{20.17c}$$

which can be written in the compact form

$$E_x = E'_x, \qquad E_y = \beta^2(E'_y + uB'_z), \qquad E_z = \beta^2(E'_z - uB'_y) \tag{20.18a}$$

$$B_x = B'_x, \qquad B_y = \beta^2\left[B'_y - \left(\frac{u}{c^2}\right)E'_z\right], \qquad B_z = \beta^2\left[B'_z + \left(\frac{u}{c^2}\right)E'_y\right] \tag{20.18b}$$

where $\beta^2 = \left(1 - \frac{u^2}{c^2}\right)^{-1}$.

Then, how do we explain this lack of self-consistency in these equations? That is, they do not transform properly into a similar set when solved with primed and unprimed quantities interchanged and u replaced by $-u$. We have to look for the reasons for this inconsistency.

This inconsistency implies that something is wrong with the underlying principles of this above analysis, i.e. either with Newtonian mechanics and Galilean relativity, or with classical electromagnetics, or with both. Historically, this was the difficulty faced by the physicists towards the end of the nineteenth century. The uncertainty and the controversy were associated with the correct formulation of the velocity of light, and was eliminated by the special (or the restricted) principle of relativity.

What we have been saying so far has been that:

> *The laws of nature are same referred to all frames of reference, which are in uniform translation with respect to one another* (i.e. there is no meaning in absolute velocity).

So, we consider, as before, two frames of reference F and F'. The coordinates of a point, in the frame F, are (x, y, z), and in the frame F' are $(x - ut, y, z)$; i.e. a particle with velocity u_1 in the x-direction with respect to the frame F, will appear to have a velocity $(u_1 - u)$ in the frame F'.

This, in fact, conflicts with the observation of constant velocity of light in both the frames of reference.

We shall not go into the historical details of this controversy regarding the velocity of light. Suffice it to say that the root of the conflict was the question of existence of an 'aether'. We have seen that Maxwell showed light to be an electromagnetic phenomenon, and that the velocity of light was given by $c = (\mu\varepsilon)^{-1/2}$—a characteristic of the medium, and presumably the velocity is measured relative to the medium. Under these circumstances, if the frame F is assumed to be stationary in the medium, the frame F' moves through the medium with a velocity u in the $+x$-direction. Then a light wave travelling in the medium with a velocity c in the

+x-direction, should appear to have a velocity (c − u) to an observer in the frame F', moving along with it. This result conflicts with our results as we obtained in Eqs. (20.18), as the factor $(1 - u^2/c^2)$ obtrudes itself. Admittedly, c is a very large velocity, and so this factor is very small; but even so its presence raises doubts about the validity of the whole theory. It is not difficult to cook Eq. (20.11) in such a way as to remove this anomaly. For example, if it could be shown that the true equation for $\mathbf{E}'$ were

$$E'_x = E_x \tag{20.19a}$$

$$E'_y = \beta(E_y - uB_z) \tag{20.19b}$$

$$E'_z = \beta(E_z + uB_y) \tag{20.19c}$$

where

$$\beta = \left(1 - \frac{u^2}{c^2}\right)^{-1/2} \tag{20.20}$$

with a similarly modified set of equations for $\mathbf{B}'$, i.e.

$$B'_x = B_x \tag{20.19d}$$

$$B'_y = \beta\left[B_y + \left(\frac{u}{c^2}\right)E_z\right] \tag{20.19e}$$

$$B'_z = \beta\left[B_z - \left(\frac{u}{c^2}\right)E_y\right] \tag{20.19f}$$

the process of solution would give

$$E_x = E'_x \tag{20.21a}$$

$$E_y = \frac{E'_y + uB'_z}{\beta\left(1 - \dfrac{u^2}{c^2}\right)} = \beta\,(E'_y + uB'_z) \tag{20.21b}$$

$$E_z = \frac{E'_z - uB'_y}{\beta\left(1 - \dfrac{u^2}{c^2}\right)} = \beta\,(E'_z - uB'_y) \tag{20.21c}$$

and so forth, so that the new equations would have the same form as Eqs. (20.19) except that the sign of u has been changed. It should be noted in passing that Eqs. (20.19) cannot be expressed directly in vector form, because the factor β does not multiply every component.

However, the mere fact that β makes the answer come true (right) does not justify us in inserting it. There must be some much more solid argument for its presence. To explain it, we need a fundamental enquiry into the relation between the physical quantities as measured in the two frames of reference; not merely of electromagnetic quantities, but of length and time, and—ultimately—of mass. For the two frames of reference having a relative motion, which is linear and uniform, this inquiry is the province of the 'Special Theory of Relativity'.

20.5 THE RELATIVITY OF SPACE AND TIME

The 'relativity' of electromagnetic measurements to a particular observer is a consequence of the elementary electromagnetic principles, but the more subtle considerations lead to a similar 'relativity' about the fundamental measures of mass, length, and time. We shall consider first the last two.

We have proved that the electromagnetic waves travel with the velocity $c = (\mu_0\varepsilon_0)^{-1/2}$ in free space. Since the field equations on which this deduction is based are relative to a particular observer, we do not expect the velocity as observed by an observer in a 'moving' frame of reference to be the same. Experiments [especially those of Michaelson and Morley (1887)] show that *the velocity of light or other electromagnetic signals is the same for all observers.*

This means a departure from the normal kinematic ideas; if a particle has a velocity (U, V, W) with respect to the frame of reference F, it is assumed to have the velocity (U – u, V, W) with respect to the moving frame F', and so it cannot have the same velocity with respect to both. In order that the light may have the same velocity c with respect to both F and F', we find it necessary to introduce a scheme wherein *the measures of length l and time t are relative to a particular observer.*

Let F and F' be the frames of reference having relative motion u, and let them coincide at $t = 0$. An *event* is defined as something happening at a specified place and time—e.g. a flash of light or a pulse of radiation occurring at a place (x, y, z) and time (t). In ordinary or classical Newtonian mechanics, an event (x, y, z, t) observed from the frame F will have the coordinates

$$x' = x - ut, \qquad y' = y, \qquad z' = z, \qquad t' = t \tag{20.22}$$

observed from the frame F'. These are linear relationships, but we have seen that they cannot satisfy the 'constant velocity of light' condition.

The path of a particle in either F or F' is a sequence of events. If a particle is acted on by no forces, it moves (we assume) with uniform velocity in a straight line with respect to either F or F'; its path may be represented by a straight line with time members uniformly spread along it. In order that this may be true, the relations between (x, y, z, t) and (x', y', z', t') must be linear.

A flash of light occurring at the origin, at the instant when the two frames F and F' coincide, will spread out so that at the time t an observer in F sees it as having reached the sphere

$$x^2 + y^2 + z^2 = c^2t^2$$

and, in F'

$$x'^2 + y'^2 + z'^2 = c^2t'^2$$

Hence the relation between (x, y, z, t) and (x', y', z', t') must be such that

$$x'^2 + y'^2 + z'^2 - c^2t'^2 = x^2 + y^2 + z^2 - c^2t^2 \tag{20.23}$$

and this combination of coordinates is said to be *invariant.*

We find it possible to satisfy [Eq. (20.23)] with $y' = y$, $z' = z$, and x', t' linear functions of x, t such that

$$x'^2 - c^2t'^2 = x^2 - c^2t^2 \tag{20.24}$$

Let those functions be (since $x = 0$, $t = 0 \rightarrow x' = 0$, $t' = 0$)

$$x' = Ax + Bt,\ t' = Cx + Dt;\ \text{then from (20.24)}$$

$$(Ax + Bt)^2 - c^2(Cx + Dt)^2 = x^2 - c^2t^2$$

whence

$$A^2 - c^2C^2 = 1, \qquad c^2D^2 - B^2 = c^2, \qquad AB - c^2CD = 0 \tag{20.25}$$

Now $x' = y' = z' = 0$ is moving with the velocity u with respect to F.

$$\therefore \qquad \frac{B}{A} = -u \tag{20.26}$$

and Eqs. (20.25) and (20.26) give four equations for A, B, C, D.

Substituting from Eq. (20.26) in Eqs. (20.25),

$$c^2C^2 = A^2 - 1 \qquad \text{and} \qquad c^2D^2 = u^2A^2 + c^2; \text{ thus}$$

$$(A^2 - 1)(u^2A^2 + c^2) - u^2A^4 = 0$$

whence (taking the positive root),

$$A = \left(1 - \frac{u^2}{c^2}\right)^{-1/2} = \beta \tag{20.27a}$$

$$B = -u\beta; \qquad C = \pm\frac{\beta u}{c^2}; \qquad D = \beta \tag{20.27b}$$

Assimilation to Eq. (20.22) requires the lower signs, and the relations are:

$$x' = \beta(x - ut), \qquad y' = y, \qquad z' = z, \qquad t' = \beta\left(t - \frac{ux}{c^2}\right) \tag{20.28}$$

This is called the *Lorentz transformation.*

20.6 CONSEQUENCES OF THE LORENTZ TRANSFORMATION

1. Events which occur simultaneously, but at different places in one frame, are not simultaneous in the other frame. This is a consequence of $t' = \beta\left(t - \frac{ux}{c^2}\right)$.
2. If the equations are solved for x', t' in terms of x and t, we get the same relation with the sign of u changed.
3. *Contraction of lengths.* Let x'_1, x'_2 be points on the x'-axis in the frame F'; and x_1, x_2 their coordinates in the other frame F. Then,

$$x'_2 - x'_1 = \beta\,(x_2 - x_1) \tag{20.29}$$

when measured at a particular time t with respect to the frame F.

Hence an observer in F sees a length divided by β, a number > 1. (This is Fitzgerald contraction.)

It should be noted that there is no contraction perpendicular to the line of motion.

4. *Dilatation of time.* Let a clock at the origin of the frame F' tick at $t' = t'_1$ and $t' = t'_2$. With respect to the frame F, we have

$$x_1 - ut_1 = 0, \qquad x_2 - ut_2 = 0, \qquad t'_1 = \beta\left(t_1 - \frac{ux_1}{c^2}\right), \qquad t'_2 = \beta\left(t_2 - \frac{ux_2}{c^2}\right)$$

i.e. $$t_1' = \beta\left(1-\frac{u^2}{c^2}\right)t_1 = \frac{t_1}{\beta}, \qquad t_2' = \frac{t_2}{\beta}$$

and the time-interval measured from the frame F is

$$t_2 - t_1 = \beta\,(t_2' - t_1') \tag{20.30}$$

—a longer time—i.e. the ticks seem to occur at a longer interval. This leads to the famous *clock paradox.*

5. *Transformation of velocities.* Let a particle have the velocity (v_x, v_y, v_z) in the frame F, and (v_x', v_y', v_z') in the frame F', where

$$v_x^2 + v_y^2 + v_z^2 = v^2 \quad \text{and} \quad v_x'^2 + v_y'^2 + v_z'^2 = v'^2$$

If the coordinates of the particle in the two frames are (x, y, z, t) and (x', y', z', t') respectively, then:

$$v_x' = \frac{dx'}{dt'} = \frac{\beta(dx - u\,dt)}{\beta\left(dt - \dfrac{u\,dx}{c^2}\right)} = \frac{v_x - u}{1 - \dfrac{uv_x}{c^2}} \tag{20.31a}$$

$$v_y' = \frac{dy'}{dt'} = \frac{dy}{\beta\left(dt - \dfrac{u\,dx}{c^2}\right)} = \frac{v_y}{\beta\left(1 - \dfrac{uv_x}{c^2}\right)} \tag{20.31b}$$

$$v_z' = \frac{dz'}{dt'} = \frac{dz}{\beta\left(dt - \dfrac{u\,dx}{c^2}\right)} = \frac{v_z}{\beta\left(1 - \dfrac{uv_x}{c^2}\right)} \tag{20.31c}$$

Similarly,

$$v_x = \frac{dx}{dt} = \frac{v_x' + u}{1 + \dfrac{v_x' u}{c^2}} \tag{20.31d}$$

$$v_y = \frac{dy}{dt} = \frac{v_y'}{\beta\left(1 + \dfrac{v_x' u}{c^2}\right)} \tag{20.31e}$$

$$v_z = \frac{dz}{dt} = \frac{v_z'}{\beta\left(1 + \dfrac{v_x' u}{c^2}\right)} \tag{20.31f}$$

It should be noted that if the two systems have a relative velocity $u = c$, and if a point in the F'-system has a velocity $v_x' = c$, then the value of v_x is still c.

By similarly differentiating these six equations with respect to t and t', we get

$$\frac{dv_x}{dt} = \left[\beta\left(1 + \frac{v_x' u}{c^2}\right)\right]^{-3} \frac{dv_x'}{dt'} \tag{20.31g}$$

and
$$\frac{dv_{y,z}}{dt} = \left[\beta\left(1+\frac{v'_x u}{c^2}\right)\right]^{-2} \frac{dv'_{y,z}}{dt'} - \frac{v'_{y,z}}{\beta^2}\frac{u}{c^2}\left(1+\frac{v'_x u}{c^2}\right)\frac{du'_x}{dt'} \tag{20.31h}$$

The corresponding expressions for $\frac{du'_x}{dt'}$ and $\frac{du'_{y,z}}{dt'}$ can be obtained by interchanging the primed and unprimed quantities and replacing +u by –u. A point to be noted is that a constant acceleration in the F'-system does not, in general, imply a constant acceleration in the F-system. Thus,

$$c^2 - v'^2 = c^2 - \frac{\beta^2(v_x-u)^2 + v_y^2 + v_z^2}{\beta^2\left(1-\frac{uv_x}{c^2}\right)^2}$$

$$= c^2 - \frac{(v_x-u)^2 + \left(1-\frac{u^2}{c^2}\right)(v_y^2+v_z^2)}{\left(1-\frac{uv_x}{c^2}\right)^2}$$

$$= \frac{\left(c^2 - 2uv_x + \frac{u^2v_x^2}{c^2}\right) - \left(v_x^2 - 2v_x u + u^2\right) - \left(1-\frac{u^2}{c^2}\right)(v_y^2+v_z^2)}{\left(1-\frac{uv_x}{c^2}\right)^2}$$

$$= \frac{\left(1-\frac{u^2}{c^2}\right)(c^2-v^2)}{\left(1-\frac{uv_x}{c^2}\right)^2}$$

i.e.
$$c^2 - v'^2 = \frac{c^2-v^2}{\beta^2\left(1-\frac{uv_x}{c^2}\right)^2} \tag{20.32}$$

$\therefore$ If v = c, then v′ = c; the 'addition' of (u, 0, 0) to (v_x, v_y, v_z), then gives the same velocity.

If we write $(1 - v^2/c^2) = (\beta_v)^{-2}$, then $(1 - v'^2/c^2) = (\beta'_v)^{-2}$
then Eq. (20.32) becomes:
$$(\beta'_v)^2 = \left(1-\frac{uv_x}{c^2}\right)\beta\beta_v \tag{20.33}$$

6. *Transformation of mass and momentum.* (a) We postulate for the mechanics of interacting particles, as a general law of nature, the conservation of mass and momentum.

In the system F', two equal perfectly elastic particles, moving with velocities v and –v along $O'x'$, collide and rebound. ('Equal' means that they have identical properties with respect to an observer at rest relative to them.)

The frame F' moves relative to the frame F, parallel to Ox with the velocity u, and it will be shown that the masses of the particles depend on their velocities. Relative to F, let the velocities before the collision be w_1 and w_2, and let the masses be m_1, m_2. Let M be the sum of the masses at the moment of collision when they are relatively at rest—at rest relative to F', but having velocity u relative to F.

Then, for F:

$$m_1 + m_2 = M \qquad \text{(conservation of mass)} \tag{20a}$$

$$m_1 w_1 + m_2 w_2 = Mu \qquad \text{(conservation of momentum)} \tag{20b}$$

Also,
$$w_1 = \frac{u + v}{1 + \frac{uv}{c^2}}, \qquad w_2 = \frac{u - v}{1 - \frac{uv}{c^2}} \tag{20c}$$

From Eqs. (20a) and (20b),

$$m_1 = \frac{M(u - w_2)}{w_1 - w_2} \qquad \text{and} \qquad m_2 = \frac{M(w_1 - u)}{w_1 - w_2}$$

$$\frac{m_1}{m_2} = \frac{u - w_2}{w_1 - u} = \frac{1 + \frac{uv}{c^2}}{1 - \frac{uv}{c^2}}, \qquad \text{by (Eq. 20c)}$$

but by Eq. (20c),

$$1 - \frac{w_1^2}{c^2} = \frac{1 - (u + v)^2}{c^2\left(1 + \frac{uv}{c^2}\right)^2}$$

$$= \frac{\left(1 - \frac{u^2}{c^2}\right)\left(1 - \frac{v^2}{c^2}\right)}{\left(1 + \frac{uv}{c^2}\right)^2}$$

Thus,

$$1 + \frac{uv}{c^2} = \left[\frac{\left(1 - \frac{u^2}{c^2}\right)\left(1 - \frac{v^2}{c^2}\right)}{1 - \frac{w_1^2}{c^2}}\right]^{1/2} \tag{20d1}$$

and

$$1 - \frac{uv}{c^2} = \left[\frac{\left(1 - \frac{u^2}{c^2}\right)\left(1 - \frac{v^2}{c^2}\right)}{1 - \frac{w_2^2}{c^2}}\right]^{1/2} \tag{20d2}$$

$$\therefore \qquad \frac{m_1}{m_2} = \left(\frac{1 - \frac{w_2^2}{c^2}}{1 - \frac{w_1^2}{c^2}} \right)^{1/2} \tag{20d3}$$

Hence, the masses of the two particles, which by hypothesis have the same mass—m_0, say, when at rest, are inversely proportional to $(1 - w^2/c^2)^{1/2}$ when moving with the velocity w. So the mass m of a particle moving with the velocity w is

$$m = \frac{m_0}{\left(1 - \frac{w^2}{c^2}\right)^{1/2}}$$

where m_0 is called its *rest-mass* or *proper-mass*. Therefore, the mass m increases with velocity. By Eq. (20a),

$$\begin{aligned} M = m_1 + m_2 &= \frac{m_0}{\left(1 - \frac{w_1^2}{c^2}\right)^{1/2}} + \frac{m_0}{\left(1 - \frac{w_2^2}{c^2}\right)^{1/2}} \\ &= \frac{m_0 \left[\left(1 + \frac{uv}{c^2}\right) + \left(1 - \frac{uv}{c^2}\right) \right]}{\sqrt{\left(1 - \frac{u^2}{c^2}\right)\left(1 - \frac{v^2}{c^2}\right)}} \qquad \text{by Eqs. (20d),} \\ &= \frac{2m_0}{\sqrt{\left(1 - \frac{u^2}{c^2}\right)\left(1 - \frac{v^2}{c^2}\right)}} \\ &> \frac{2m_0}{\sqrt{\left(1 - \frac{u^2}{c^2}\right)}}, \quad \text{since } \left(1 - \frac{v^2}{c^2}\right) < 1, \quad \text{when } v \neq 0. \end{aligned}$$

But M is the total mass at the instant of collision, when the particles are relatively at rest—the instant of greatest deformation. u is their common velocity at this instant (with respect to F); $2m_0$ is their total (undeformed) rest-mass.

Hence the total mass M at the instant of greatest deformation is greater than the mass with respect to F as calculated from the total undeformed rest-mass. This suggests that the mass must be ascribed to the potential energy of the elastic deformation, if the laws of conservation of mass and momentum are to be preserved.

(b) We consider next a simple hypothetical experiment devised by Tolman.

As before we consider two frames of reference F and F' where F' moves with a constant velocity u in the x-direction with respect to the frame F (also they have a common x-axis). At the instant of time when the origins of the two frames are coincident, the frame F' projects a sphere with a velocity v from B' toward O', and the frame F projects one from A parallel to y with a velocity v. The lengths OA and OB are so chosen that these two spheres collide when their centres are aligned in the y-direction. The collisions as observed by F and F' are shown in Figure 20.4(a).

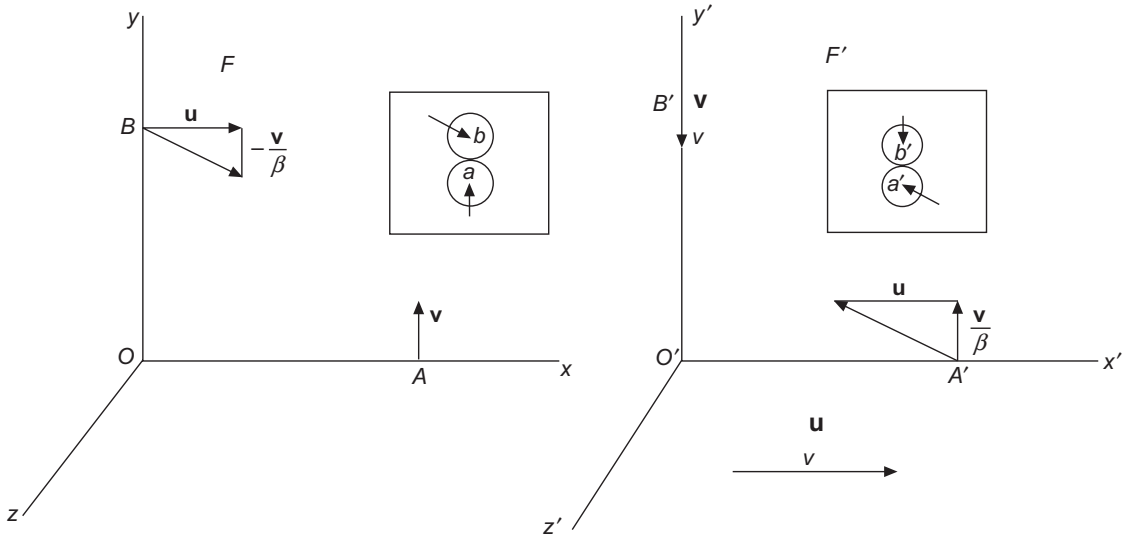


Figure 20.4(a) Experiment with colliding spheres.

The masses of the spheres have been so chosen that when they are at rest with respect to **any** observer, both have (for the observer), the mass m_0. It is assumed that the mass is a function of the magnitude of the velocity, so that

$$m = f(\mathrm{v}^2) \qquad \text{and} \qquad m_0 = f(0) \tag{20.d4}$$

From the formulae of last section,

For F	For F'
$\mathrm{v}_{ax} = 0, \quad \mathrm{v}_{ay} = \mathrm{v}$	$\mathrm{v}'_{ax} = -\mathrm{u} \quad \mathrm{v}'_{ay} = \dfrac{\mathrm{v}}{\beta}$
$\mathrm{v}_{bx} = \mathrm{u} \quad \mathrm{v}_{by} = -\dfrac{\mathrm{v}}{\beta}$	$\mathrm{v}'_{bx} = 0 \quad \mathrm{v}'_{by} = -\mathrm{v}$

(20.d5)

Let the velocities after the impact be denoted by bars on the top.

Neither of the observers in the two frames will see any transfer of the x-component of the momentum (as a consequence of the impact of smooth spheres),

Hence F observes that $\quad m_b \mathrm{u} = \bar{m}_b \mathrm{u},$

or

$$\mathrm{u} f\left(\mathrm{u}^2 + \frac{\mathrm{v}^2}{\beta^2}\right) = \mathrm{u} f\left(\mathrm{u}^2 + \frac{\bar{\mathrm{v}}_b^2}{\beta^2}\right)$$

which requires $\quad \bar{\mathrm{v}}_b = -\mathrm{v} \quad$ and $\quad \therefore \quad \bar{m}_b = m_b \tag{20.d6}$

The –ve sign has been chosen, since it is to be correct for small velocities where $\beta = 1$.

For the y-component, the observation of F is

$$\mathrm{v}f(\mathrm{v}^2) - \frac{m_b \mathrm{v}}{\beta} = \bar{\mathrm{v}}_a f(\bar{\mathrm{v}}_a^2) - \frac{\bar{m}_b\, \bar{\mathrm{v}}_b}{\beta}$$

Since we saw above that $\bar{m}_b = m_b$ and also the known result when u = 0, this equation can be satisfied, only if $\bar{\mathrm{v}}_a = -\mathrm{v}$, and so the above equation becomes:

$$f(\mathrm{v}^2) = m_a = \frac{m_b}{\beta} \tag{20.d7}$$

If the velocity of projection v becomes very small, then

$$m_b = f\left(\mathrm{u}^2 + \frac{\mathrm{v}^2}{\beta^2}\right) \to f(\mathrm{u}^2) \quad \text{and} \quad m_a = f(\mathrm{v}^2) \to f(0) = m_0 \tag{20.d8}$$

Thus the mass of an object in motion having a velocity u with respect to a given observer appears to him to be increased by a factor

$$\beta = \left\{1 - \left(\frac{\mathrm{u}}{\mathrm{c}}\right)^2\right\}^{-1/2} \tag{20.d9}$$

over its mass at rest.

At this stage, two new symbols are introduced, i.e. β_1 and β_1' which are defined as:

$$\beta_1 = \left(1 - \frac{\mathrm{v}_x^2 + \mathrm{v}_y^2 + \mathrm{v}_z^2}{\mathrm{c}^2}\right)^{-1/2} \tag{20.d10}$$

$$\beta_1' = \left(1 - \frac{\mathrm{v}_x'^2 + \mathrm{v}_y'^2 + \mathrm{v}_z'^2}{\mathrm{c}^2}\right)^{-1/2} \tag{20.d11}$$

where v_x, v_y, v_z are the components of the velocity of a particle as seen by an observer in the frame F, and v_x', v_y', v_z' are the components of the velocity of the same particle as seen by the observer in the frame F'.

Using Eqs. 20.31(a) to (f), we get

$$\beta_1 = \beta\beta'_1\left(1 + \frac{\mathrm{v}_x' \mathrm{u}}{\mathrm{c}^2}\right) \tag{20.d12}$$

and

$$\beta'_1 = \beta\beta_1\left(1 - \frac{\mathrm{v}_x \mathrm{u}}{\mathrm{c}^2}\right) \tag{20.d13}$$

7. *Transformation of charge density amd current density.* (If ρ_C and **J** are the values measured from the frame F, what are the corresponding values measured from the frame F'?)

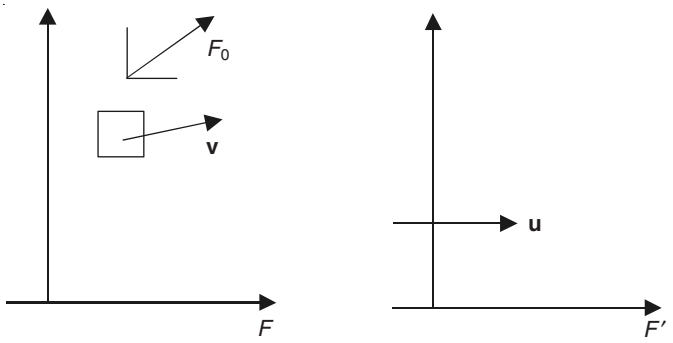


Figure 20.5 Charge density and current in the two frames of reference F and F'.

We consider a volume element δv with respect to the frame F, containing $(n_v \delta v)$ charges, all moving with the same velocity $\mathbf{v}$ $(= v_x, v_y, v_z)$. Let F_0 be a frame of reference moving with the charges; and if an observer in F_0 sees $(n_0 \delta v_0)$ charges occupying the volume δv_0, then the observer in the frame F sees:

$$\delta v = \delta v_0 \sqrt{\left(1 - \frac{v^2}{c^2}\right)}, \quad \text{and since } (n_v \delta v) = (n_0 \delta v_0),$$

$$n_v = \frac{n_0}{\sqrt{\left(1 - \frac{v^2}{c^2}\right)}} \tag{20.34}$$

If each charge is q (assumed absolute), we have

$$\rho_C = \Sigma (q n_v) \tag{20.35}$$

and

$$J_x = \Sigma (q n_v v_x), \quad \text{and similarly } J_y \text{ and } J_z, \text{ so that:}$$

$$\mathbf{J} = \Sigma (q n_v \mathbf{v}) \tag{20.36}$$

where the summation is over all the velocities.

Similarly, an observer in the frame F' sees the density:

$$n'_v = \frac{n_0}{\sqrt{\left(1 - \frac{v'^2}{c^2}\right)}} \tag{20.37}$$

$$\rho'_C = \Sigma (q n'_v) \tag{20.38}$$

$$\mathbf{J}' = \Sigma (q n'_v \mathbf{v}') \tag{20.39}$$

The components of $\mathbf{v}'$ are given by Eq. (20.31), and in the notation of Eq. (20.33), Eqs. (20.34) and (20.37) are to be written in the form:

$$n_v = \beta_v n_0, \quad n'_v = \beta'_v n_0 \tag{20.40}$$

$$\therefore \qquad n'_v = \left(\frac{\beta'_v}{\beta_v}\right) n_v = \beta \left(1 - \frac{u v_x}{c^2}\right) n_v \quad \text{[from Eq. (20.33)]} \tag{20.41}$$

Thus,

$$\rho'_{\mathrm{C}} = \Sigma\,(qn'_{\mathrm{v}}) = \Sigma\left[q\beta\left(1-\frac{\mathrm{u}\mathrm{v}_x}{\mathrm{c}^2}\right)n_{\mathrm{v}}\right]$$

$$= \beta\left(\rho_{\mathrm{C}} - \frac{\mathrm{u}J_x}{\mathrm{c}^2}\right) \tag{20.42}$$

$$J'_x = \Sigma\,(qn'_{\mathrm{v}}\mathrm{v}'_x) = \Sigma\left[q\beta\left(1-\frac{\mathrm{u}\mathrm{v}_x}{\mathrm{c}^2}\right)\left(\frac{\mathrm{v}_x-\mathrm{u}}{1-\frac{\mathrm{u}\mathrm{v}_x}{\mathrm{c}^2}}\right)n_{\mathrm{v}}\right]$$

$$= \beta\,(J_x - \rho_{\mathrm{C}}\mathrm{u}) \tag{20.43}$$

$$J'_y = \Sigma\,(qn'_{\mathrm{v}}\mathrm{v}'_y) = \Sigma\left[q\beta\left(1-\frac{\mathrm{u}\mathrm{v}_x}{\mathrm{c}^2}\right)\frac{\mathrm{v}_y n_{\mathrm{v}}}{\beta\left(1-\frac{\mathrm{u}\mathrm{v}_x}{\mathrm{c}^2}\right)}\right]$$

$$= J_y \tag{20.44}$$

and

$$J'_z = J_z \tag{20.45}$$

These are similar to the Lorentz transformation of t, x, y, z.

When $\beta \to 1$, they tend to vector equations:

$$\rho'_{\mathrm{C}} = \rho_{\mathrm{C}} - \left(\frac{1}{\mathrm{c}^2}\right)(\mathbf{u}\cdot\mathbf{J}) \tag{20.46}$$

$$\mathbf{J}' = \mathbf{J} - \rho_{\mathrm{C}}\mathbf{u} \tag{20.47}$$

8. *Moving conductors.* The charge and the current density transformations derived in Eqs. (20.42)–(20.45) can be expressed in the compact matrix form as

$$\begin{bmatrix} J_x \\ J_y \\ J_z \\ j\mathrm{c}\rho_{\mathrm{C}} \end{bmatrix} = \begin{bmatrix} \beta & 0 & 0 & -j(\beta\mathrm{u}/\mathrm{c}) \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ j(\beta\mathrm{u}/\mathrm{c}) & 0 & 0 & \beta \end{bmatrix} \begin{bmatrix} J'_x \\ J'_y \\ J'_z \\ j\mathrm{c}\rho'_{\mathrm{C}} \end{bmatrix} \tag{20.48}$$

(a) CHARGED BODY NOT CARRYING CONDUCTION CURRENT

If $J'_x = J'_y = J'_z = 0$, then we get $J_x = (\beta\mathrm{u}\rho'_{\mathrm{C}})$, $\rho_{\mathrm{C}} = \beta\rho'_{\mathrm{C}}$.

The latter is increased due to the Fitzgerald contraction. The former, equivalent to $J_x = (\mathrm{u}\rho_{\mathrm{C}})$, is a convection current.

(b) CURRENT-CARRYING BODY NOT CHARGED

If $\rho'_{\mathrm{C}} = 0$, we get $J_x = \beta J'_x$, $\quad J_y = J'_y$, $\quad J_z = J'_z$, $\quad \rho_{\mathrm{C}} = \left(\frac{\beta\mathrm{u}}{\mathrm{c}^2}\right)J'_x$.

$\mathbf{J}'$ is entirely the conduction current (the movement of equal +ve and −ve charges). Ohm's law ($\mathbf{E}' = \rho'\mathbf{J}'$) is obeyed.
(Note that ρ' is the conductivity of the conductor, whereas ρ_C and ρ'_C are the charge densities.)

From the fixed frame, the +ve and −ve charges no longer appear equal, and the current (density vector) $\mathbf{J}$ is part conduction, and part convection; and in fact:

$$\mathbf{J} = \mathbf{J} + \rho_C \mathbf{u} \tag{20.49}$$

Ohm's law applies to conduction current only. We thus have

$$J_{Cx} = J_x - \rho_C u = \beta J'_x - \left(\frac{\beta u^2}{c^2}\right) J'_x = \frac{J'_x}{\beta}$$

Also, $J_{Cy} = J_y = J'_y$, $J_{Cz} = J'_z$.

Let the electric and the magnetic fields in the fixed frame F be $\mathbf{E}$, $\mathbf{B}$ respectively. The effective electric force driving the current through the conductor is $[\mathbf{E} + (\mathbf{u} \times \mathbf{B})]$; so if the effective resistivities are ρ_x, ρ_y, ρ_z, then

$$E_x = \rho_x J_{Cx}, \qquad E_y - uB_z = \rho_y J_{Cy}, \qquad E_z + uB_y = \rho_z J_{Cz}$$

But $E_x = E'_x = \rho' J'_x$, $\qquad E_y - uB_z = \dfrac{E'_y}{\beta} = \rho' \dfrac{J'_y}{\beta}$, $\qquad E_z + uB_y = \rho' \dfrac{J'_z}{\beta}$.

$\therefore$ Substituting for J_{Cx}, J_{Cy}, J_{Cz}, we get

$$\rho' J'_x = \frac{\rho_x J'_x}{\beta} \quad \text{or} \quad \rho_x = \beta\rho' \tag{20.50a}$$

$$\frac{\rho' J'_y}{\beta} = \rho_y J'_y \quad \text{or} \quad \rho_y = \frac{\rho'}{\beta} \tag{20.50b}$$

$$\frac{\rho' J'_z}{\beta} = \rho_z J'_z \quad \text{or} \quad \rho_z = \frac{\rho'}{\beta} \tag{20.50c}$$

Thus ρ (= the resistivity) appears to have increased in the direction of motion, and reduced in the perpendicular directions.

20.6.1 The Transformation Equations for Force

We define the force as the rate of change of momentum (and not as the product of mass by acceleration, which can then change the nature of these equations). Then,

$$\mathbf{F} = \frac{d}{dt}(m\mathbf{v}) = m_0 \frac{d}{dt}(\beta_1 \mathbf{v}) \tag{20.50d}$$

When the components of this equation are written down, it is seen that with this definition, the force and the acceleration, in general, are not in the same direction. Carrying out the differentiation in the above equation,

$$\mathbf{F} = m\frac{d\mathbf{v}}{dt} + \mathbf{v}\frac{dm}{dt} = m_0\beta_1 \frac{d\mathbf{v}}{dt} + m_0\beta_1^3 \, \mathbf{v}\frac{v}{c^2}\frac{dv}{dt}$$

$$\left[\frac{dm}{dt}=\frac{d}{dt}(m_0\,\beta_1),\ \beta_1=\left(1-\frac{v_x^2+v_y^2+v_z^2}{c^2}\right)^{-1/2}=\left(1-\frac{v^2}{c^2}\right)^{-1/2}\right.$$

$$\left.=m_0\left(\frac{d\beta_1}{dt}\right)=m_0\left\{-\frac{1}{2}\left(1+\frac{v^2}{c^2}\right)^{-\frac{1}{2}-1}\left(-\frac{2v}{c^2}\right)\frac{dv}{dt}\right\}=m_0\beta_1^3\,\frac{v}{c^2}\cdot\frac{dv}{dt}\right]$$

$$\therefore\qquad \mathbf{F}=m_0\beta_1^3\left[\left(1-\frac{v^2}{c^2}\right)\frac{d\mathbf{v}}{dt}+\frac{v\mathbf{v}}{c^2}\frac{dv}{dt}\right] \tag{20.50e}$$

If the force applied is in the direction $\mathbf{v}_1$ of v ($\mathbf{v}_1$ = unit vector in the direction of $\mathbf{v}$), then

$$\mathbf{v}=\mathbf{v}_1 v \quad \text{and} \quad \frac{d\mathbf{v}}{dt}=\mathbf{v}_1\frac{dv}{dt}$$

then
$$\mathbf{F}_l=m_0\beta_1^3\frac{d\mathbf{v}}{dt}=m_l\frac{d\mathbf{v}}{dt} \tag{20.50f}$$

where $\mathbf{F}_l$ is the longitudinal component of the force, and $m_l=m_0\beta_1^3$ = longitudinal mass of the particle.

If $\frac{dv}{dt}=0$, i.e. the velocity changes in direction, but not in magnitude, then Eq. (20.50e) shows that the force is again in the direction of acceleration (which is perpendicular to $\mathbf{v}$) and hence,

$$\mathbf{F}_t=m_0\beta_1\frac{d\mathbf{v}}{dt}=m_t\frac{d\mathbf{v}}{dt} \tag{20.50g}$$

where $\mathbf{F}_t$ is the transverse component of the force, and $m_t=m_0\beta_1$ = transverse mass of the particle.

The components of the force $\mathbf{F}$ can be obtained as:

$$F_x=m_0\frac{d}{dt}(\beta_1 v_x) \qquad \text{[from Eq. (20.50d)]}$$

$$=m_0\left(1+\frac{v_x' u}{c^2}\right)^{-1}\frac{d}{dt'}\{\beta_1'(v_x'+u)\}$$

$$=F_x'+m_0 u\left(1+\frac{v_x' u}{c^2}\right)^{-1}\left[\left(1-\frac{v_x'^2}{c^2}\right)\frac{d}{dt'}\beta_1'-\frac{\beta_1' v_x'}{c^2}\frac{d}{dt'}v_x'\right] \tag{20.50h}$$

and so on.

20.6.2 Energy of a Charged Moving Particle

The energy given to a charged particle by the action of a force is defined as the work done on it.

If in an infinitesimal interval of time dt, a particle moves a distance $\mathbf{dr} = \mathbf{v}dt$, then the work done on it,

$$dW = \mathbf{F}\cdot\mathbf{dr} = F_l dr = m_0\beta_1^3 \frac{dv}{dt} dr \qquad \text{[from Eq. (20.50f)]}$$

$$= m_0\beta_1^3 \, dv \tag{20.50j}$$

Substituting for β_1 and integrating from 0 to v, we get

$$W = \int_0^v dW = \int_0^v m_0\beta_1^3 \, dv$$

$$= (m_t - m_0)c^2 = (\beta_1 - 1)m_0c^2 \tag{20.50k}$$

i.e. an energy increase is equivalent to a mass increase.

20.7 THE PRINCIPLE OF SPECIAL RELATIVITY

The theory of special relativity is concerned with the frames of reference for which the relative velocity is uniform and in a straight line. The *principle of special relativity* states that:

> *The laws of nature are the same referred to all frames of reference in uniform translation relative to each other.*

As a consequence of applying this to the law of conservation of momentum, it has been shown that a particle whose mass is m_0 with respect to a frame in which it is at rest, has a mass $m = \beta m_0$ with respect to a frame in which it has a velocity u. Thus, *moving bodies become heavier.* This *relativistic increase of mass* must be taken into account in designing high-energy particle accelerators.

Note that all these effects tend to a limit, as u $\rightarrow$ c—moving rod contracts to zero, moving clock goes infinitely slowly, moving mass becomes infinite.

The velocity c of light in free space is deduced from the Maxwell's equations. Since this velocity applies to both the frames, we expect the Maxwell's equations to transform into themselves, or to be *invariant under a Lorentz transformation.* We shall verify this later.

20.8 FIRST ORDER RELATIVISTIC EFFECTS IN ELECTROMAGNETISM

In the exact equations relating $\mathbf{E}'$, $\mathbf{B}'$ to $\mathbf{E}$, $\mathbf{B}$, the factor β appears in the equations for the transverse components only. This means that they cannot be written in the vector form without the use of a highly artificial 'transverse stretching operator'. However if we write $\beta = 1$ [so neglecting the (u^2/c^2) term], we get a *first-order relativity* which cannot be entirely self-consistent, but this nevertheless throws light on a number of electromagnetic phenomena. In this (simplified, approximate, first-order) theory, $\mathbf{E}'$, and $\mathbf{B}'$ are related to $\mathbf{E}$ and $\mathbf{B}$ by:

$$\mathbf{E}' = \mathbf{E} + (\mathbf{u} \times \mathbf{B}) \tag{20.51a}$$

$$\mathbf{B}' = \mathbf{B} - \left(\frac{1}{c^2}\right)(\mathbf{u} \times \mathbf{E}) \tag{20.51b}$$

There is no loss of generality, if we take the direction of the relative velocity of the frames $\mathbf{u}$ as the x-axis; so that the coordinates are now related by the classical equations:

$$x' = x - \mathrm{u}t, \qquad y' = y, \qquad z' = z, \qquad t' = t$$

A charge density ρ_C at rest in the frame F is observed as moving with a velocity $(-\mathbf{u})$ in the frame F'; thereby constituting a convection current density $(-\rho_C \mathrm{u})$; Thus the current densities $\mathbf{J}$ and $\mathbf{J}'$ are related by the equation:

$$\mathbf{J}' = \mathbf{J} - \rho_C \mathbf{u} \tag{20.52}$$

It is not so easily obvious physically that a current density in the frame F gives rise to a charge density when viewed from the moving frame F'; this however is a consequence of the first equation of Eq. (20.51a), as we shall see. Let there be a steady current density $\mathbf{J}$ in the frame F; this creates a magnetic field which is such that it satisfies the equation:

$$\text{curl } \mathbf{H} = \nabla \times \mathbf{H} = \mathbf{J}$$

Let the currents and the charges be in free space; then the first equation of Eq. (20.51), when multiplied by ε_0, can be written in the form:

$$\mathbf{D}' = \mathbf{D} + \left(\frac{1}{\mathrm{c}^2}\right)(\mathbf{u} \times \mathbf{H}) \tag{20.53}$$

Now, in the frame F',

$$\rho'_C = \frac{\partial \mathbf{D}'_x}{\partial x'} + \frac{\partial \mathbf{D}'_y}{\partial y'} + \frac{\partial \mathbf{D}'_z}{\partial z'}$$

$$= \frac{\partial \mathbf{D}'_x}{\partial x} + \frac{\partial \mathbf{D}'_y}{\partial y} + \frac{\partial \mathbf{D}'_z}{\partial z}$$

since t is independent of the space coordinates.

$$\text{Also, } D'_x = D_x, \qquad D'_y = D_y - \left(\frac{\mathrm{u}}{\mathrm{c}^2}\right)H_z, \qquad D'_z = D_z + \left(\frac{\mathrm{u}}{\mathrm{c}^2}\right)H_y,$$

so that

$$\rho'_C = \left(\frac{\partial \mathbf{D}_x}{\partial x} + \frac{\partial \mathbf{D}_y}{\partial y} + \frac{\partial \mathbf{D}_z}{\partial z}\right) - \left(\frac{\mathrm{u}}{\mathrm{c}^2}\right)\left(\frac{\partial \mathbf{H}_z}{\partial y} - \frac{\partial \mathbf{H}_y}{\partial z}\right)$$

$$= \rho_C - \left(\frac{\mathrm{u}}{\mathrm{c}^2}\right)J_x \tag{20.54}$$

This is generalized as:

$$\rho'_C = \rho_C - \left(\frac{1}{\mathrm{c}^2}\right)(\mathbf{u} \cdot \mathbf{J}) \tag{20.55}$$

A physical explanation of the last term of Eq. (20.55) can be offered in terms of the Fitzgerald contraction. Let us imagine a conductor of unit cross-section, carrying a longitudinal current J. and let this consist of positive charges moving to the right (as shown in Figure 20.6) with velocity u, and stationary negative charges. The charge density viewed from a frame F fixed in the conductor, must be $(\pm J/\mathrm{u})$. If there are N charges, each $(-q)$, in each metre of the conductor, then: $Nq = J/\mathrm{u}$.

The N positive charges $(+q)$ must also occupy a metre when viewed from the frame F, and hence β metres when viewed from the moving frame F'; because their (i.e. the moving charges) space undergoes a Fitzgerald contraction when viewed from any frame other than that moving with the charges.

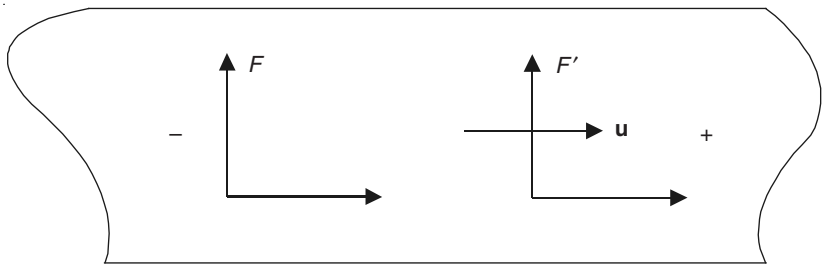


Figure 20.6 Section of a conductor of unit cross-section.

Let us now move to the frame F', and look back at the negative charges, and we observe that their space has undergone a contraction as is now $(1/\beta)$ metres. The negative charges have now become denser, and

$$\text{the apparent charge density is} = -\left(\frac{J}{\mathrm{u}}\right)\left(\beta - \frac{1}{\beta}\right) = -\left(\frac{\beta J}{\mathrm{u}}\right)\left(\frac{\mathrm{u}^2}{\mathrm{c}^2}\right)$$

To the first-order approximation, we now put $\beta = 1$. and thereby obtain a charge density (as viewed from the frame F') of $= -(\mathrm{u}J/\mathrm{c}^2)$. This is identical with the term $-(1/\mathrm{c}^2)(\mathbf{u} \cdot \mathbf{J})$ in Eq. (20.55).

This explanation given above is not a completely satisfactory one. This is because of the fact that the second-order Fitzgerald contraction has to be invoked to explain Eq. (20.55) which is derived from the first-order Eq. (20.51), and thus does indicate the difficulty of defining the border-line between the first- and the second-order effects in such a way that the first-order effects form a closed and self-consistent set.

It is not quite clear that the last term is of such an order as to be retained; but it is essential in order to be consistent with the field equations. It represents the source of the electric part, as viewed from the frame F', of the electromagnetic field which corresponds with a purely magnetic field as seen from the frame F.

In free space, Eqs. (20.51) may be rewritten as

$$\mathbf{D}' = \mathbf{D} + \left(\frac{1}{\mathrm{c}^2}\right)(\mathbf{u} \times \mathbf{H}) \tag{20.56a}$$

$$\mathbf{H}' = \mathbf{H} - (\mathbf{u} \times \mathbf{D}) \tag{20.56b}$$

Through the Maxwell's equations,

$$\text{div } \mathbf{D} = \nabla \cdot \mathbf{D} = \rho_{\mathrm{C}}$$

and

$$\text{curl } \mathbf{H} = \nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}$$

these above equations (20.56) are related to the charge and the current densities. If the transformations (20.52) and (20.55) hold good in all media—and it seems likely that the transformations of charge and current will be independent of the medium in which they are found—it follows that Eqs. (20.56) must apply to all media, not merely to free space.

20.9 THE FIELD OF A MOVING CHARGE

Let a charge q, momentarily at the origin of the frame F, be moving with respect to F with a velocity u, and let the frame F' move with the charge. Then, since the field in F' is electrostatic,

$$\mathbf{E}' = -\left(\frac{q}{4\pi\varepsilon_0}\right)\text{grad}\left(\frac{1}{r'}\right) \tag{20.57a}$$

$$\mathbf{B}' = 0 \tag{20.57b}$$

Transferring to the frame F, and noting that $r = r'$ at the instant stated, we get

$$\mathbf{E} = -\left(\frac{q}{4\pi\varepsilon_0}\right)\text{grad}\left(\frac{1}{r'}\right) \tag{20.58}$$

$$\mathbf{B} = \left(\frac{1}{c^2}\right)(\mathbf{u}\times\mathbf{E}') = -\left(\frac{q}{4\pi\varepsilon_0 c^2}\right)\left[\mathbf{u}\times\text{grad}\left(\frac{1}{r}\right)\right]$$

or

$$\mathbf{B} = \frac{\left(\frac{\mu_0 q}{4\pi}\right)(\mathbf{u}\times\mathbf{r}_1)}{r^2} \tag{20.59}$$

where $\mathbf{r}_1$ is the unit vector in the direction $\mathbf{r}$, because:

$$\text{grad}\left(\frac{1}{r}\right) = -\frac{\mathbf{r}_1}{r^2}$$

This magnetic vector is perpendicular to $\mathbf{u}$ and to $\mathbf{r}$, the rotation from $\mathbf{u}$ to $\mathbf{r}$ being right-handed about it; and also its magnitude is

$$\frac{\mu_0 q\,\text{u}\sin\theta}{4\pi r^2}$$

where θ is the angle between $\mathbf{u}$ and $\mathbf{r}$. This is in strict accord with Biot–Savart.

If we have an element δs of a circuit of wire of cross-section S (Figure 20.7), N charges per unit volume moving with velocity u, the above formula gives

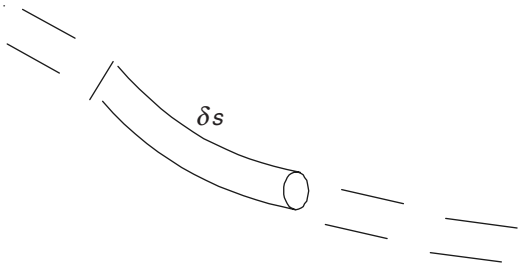


Figure 20.7 A circuit element δs.

$$\mathbf{B} = \frac{\left(\frac{\mu_0 NSq\delta s\mathrm{u}}{4\pi}\right)(\mathbf{u}\times\mathbf{r}_1)}{r^2}$$

$$= \frac{\mu_0 i(\delta\mathbf{s}\times\mathbf{r}_1)}{4\pi r^2}, \qquad \text{à la Biot–Savart} \tag{20.60}$$

20.10 THE FORCE ON A MOVING CHARGE (RELATIVISTIC CALCULATION)

If a particle is momentarily at rest in a frame F' and has the force $\mathbf{P}' = (P'_x, P'_y, P'_z)$ upon it, the components of the force measured from the frame F can be shown to be

$$P_x = P'_x, \qquad P_y = \frac{P'_y}{\beta}, \qquad P_z = \frac{P'_z}{\beta} \tag{20.61}$$

Let us consider now a charge q at rest in the frame F', so that

$$\mathbf{P}' = q\mathbf{E}'$$

or

$$P'_x = qE'_x, \qquad P'_y = qE'_y, \qquad P'_z = qE'_z$$

Transferring to the frame F, we get

$$P_x = qE_x, \qquad \beta P_y = \beta q(E_y - \mathrm{u}B_z), \qquad \beta P_z = \beta q(E_z + \mathrm{u}B_y)$$

The β's cancel, and we are left with the three equations which are equivalent to the vector equation

$$\mathbf{P} = q[\mathbf{E} + (\mathbf{u}\times\mathbf{B})] \tag{20.62}$$

which is of course the *Lorentz formula*, which is thus seen to hold good for any velocity. Both parts of it are known from elementary considerations, but the deduction of the magnetic half from the force on a current-carrying wire is of doubtful validity, and the relativistic proof is the sound one.

Note that the magnetic force $[q(\mathbf{u}\times\mathbf{B})]$ cannot change the energy of the particle.

20.11 THE BETATRON

We now consider the application of the methods discussed, to the problem of a practical device, such as Betatron.

Betatron is a particle accelerator in which the electrons are accelerated to high energies by having them move in a circular path of constant radius and at the same time, increasing the magnetic flux through the circular orbit so that the electrons acquire additional energy during each revolution. In the operation of the Betatron, the electrons from a heated filament are injected into the circular or doughnut-shaped tube by applying a difference of potential between the filament and a plate P, whilst the electrons are focused by a grid, placed in between the two. The alternating magnetic field is applied parallel to the axis of the tube. As a result, two effects are produced, i.e. an emf is produced in the electron orbit by the changing magnetic field which

gives the electrons additional energy; and secondly, a radial force is produced by the action of the magnetic field whose direction is perpendicular to the electron velocity. thus keeping the electron moving in a circular orbit. The magnetic flux through the orbit has to be chosen in such a way that the electrons will move in a stable orbit of fixed radius R. The electrons make several hundred thousand revolutions in this circular path while the alternating magnetic field is increasing from zero to a maximum in a quarter of a cycle.

The electron accelerator is like a transformer with its secondary replaced by an evacuated glass tube containing an electron source. The core is shaped as shown in the Figure 20.8, to provide at the electron orbit a field to keep the electrons in circular motion.

The equation of motion in orbit (neglecting relativity) is

$$m\left(\frac{du}{dt}\right)=\left(\frac{e}{2\pi R}\right)\left(\frac{d\Phi}{dt}\right)$$

So an electron starting with negligible velocity when $\Phi = 0$, subsequently has the velocity:

$$\mathrm{u} = \left(\frac{e}{2\pi Rm}\right)\Phi \tag{20.63}$$

Let the flux density at the orbit be B. Then, the inward radial force = $e\mathrm{u}B$. This must equal $m\mathrm{u}^2/R$.

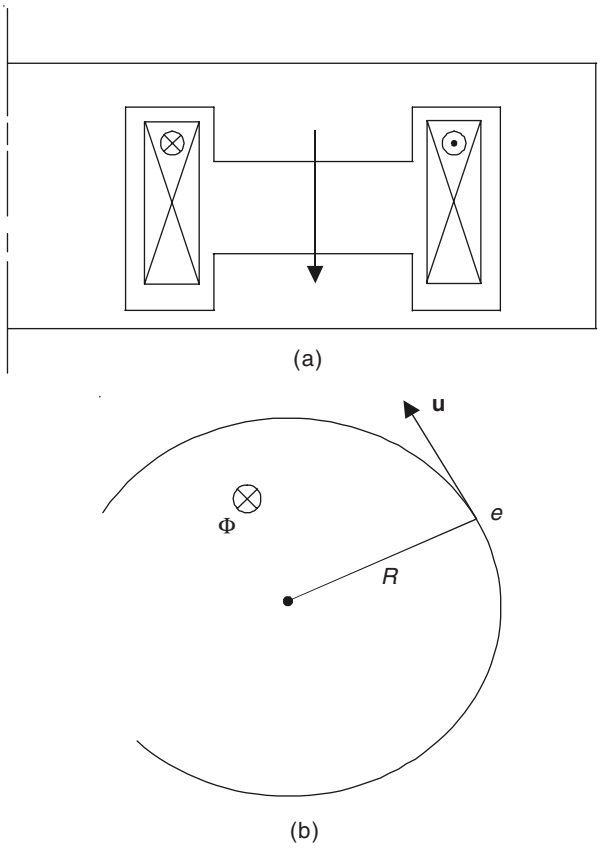


Figure 20.8 (a) Section of a Betatron and (b) the orbital path of electrons in a Betatron.

$$\therefore \qquad B = \frac{m\mathrm{u}}{eR} \tag{20.64}$$

$$\therefore \qquad B = \frac{\Phi}{2\pi R^2} = \left(\frac{1}{2}\right)(\text{Mean flux density within the orbit}) \tag{20.65}$$

The whole acceleration takes place in a quarter cycle of the supply voltage, during which time the electrons my describe 10^5 orbits. The relativistic increase of mass is used to get them out; they spiral outwards till they hit a target.

20.12 MAGNETIC FOCUSSING OF ELECTRON BEAM BY AXIAL FIELD

Let the electrons be emitted with axial and radial velocities u_0, v_0 respectively (as shown in Figure 20.9), in an axial magnetic field B, and let the subsequent velocities be u, v. Only v cooperates with B to produce a force (evB) perpendicular to the component of velocity at right angles to the axis; it should be noted that *neither component changes if there is no electric field.* In absence of the electric field E, $(u^2 + v^2)$ is constant, since B can produce no force in its direction; therefore, for all time, $u = u_0$, and $v = v_0$.

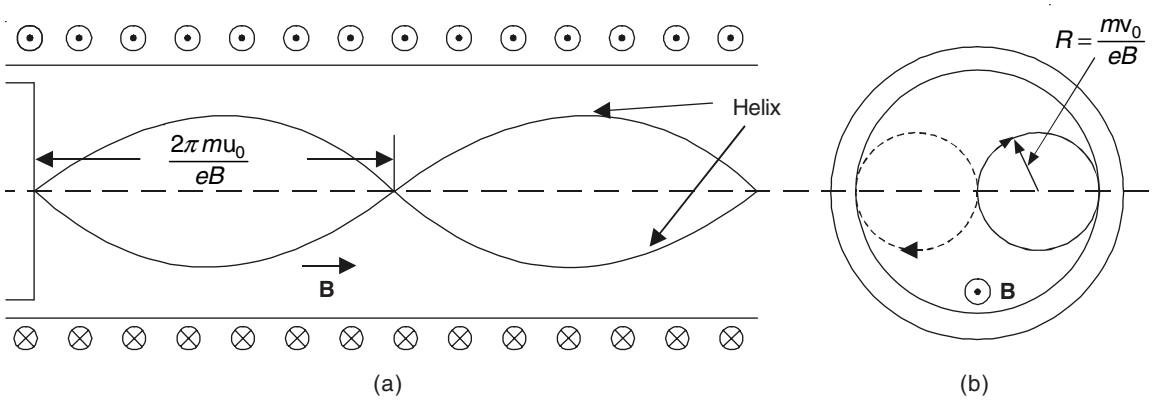


Figure 20.9 Magnetic focussing of an electron beam: (a) side view and (b) end view.

Thus the end-wise view of the path is a circle of radius $[mv_0/(eB)]$, described in the time = $[2\pi m/(eB)] = [(3.57 \times 10^{-11})/B]$ sec.

Thus the electron path is a helix, which crosses the axis of the tube at a distance equal to $[2\pi m u_0/(eB)]$ from the source. Since this does not depend on the velocity component v_0, all the electrons are brought back to this point. In practice, the electrons are likely to have equal energies, i.e. equal values of $(u_0^2 + v_0^2)$; thus the condition that all electrons shall have equal axial velocities is that (v_0/u_0) shall be small.

20.13 FOUR-DIMENSIONAL SPACE-TIME

We have seen that an event is designated by four numbers—3 space-coordinates for its position. and 1 time-coordinate for its instant of occurrance. This leads us to consider a four-dimensional

coordinate system, i.e. x, y, z, $\omega = ct$; by dropping z, we can picture the rest as a three-dimensional system. A cross-section of this system in the plane $Ox\omega$ is shown below in Figure 20.10.

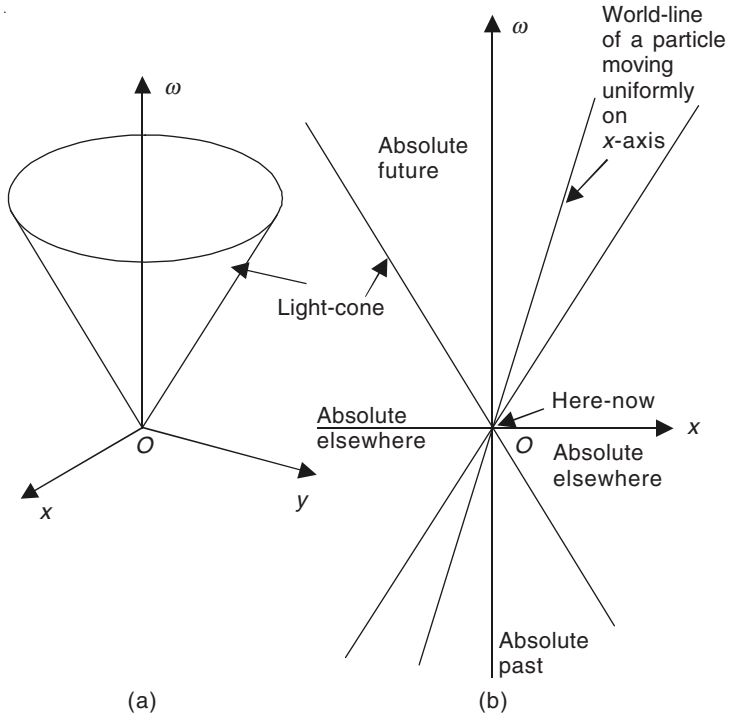


Figure 20.10 (a) Three-dimensional representation and (b) two-dimensional section of the four-dimensional space-time.

A pulse of light at $t = 0$ at O (the origin) spreads out as

$$x^2 + y^2 + z^2 + \omega^2 = 0 \tag{20.66}$$

if z is dropped, this is the equation of a cone; and so in four-dimensions we may call it a *hyper-cone*; it is also known as *light-cone*. In this four-space, each particle, moving in space and durable in time, is represented by a line which is called its *world-line*. In particular, a particle moving uniformly along the x-axis is represented by a straight world-line which must lie (since its velocity is necessarily < c) within the light-cone. If the point O (in Figure 20.10) represents *here-now*, then the line points from the past to the future. It can be shown that all the events in the upper-half of the cone are future, and all the events in the lower-half of the cone are past, with respect to O, in all reference systems; and so these regions are labeled as *absolute future* and *absolute past.*

To go from O to the region outside the cone, a particle would need to travel faster than light; this being impossible, these points are *absolute elsewhere* with respect to O. Two events can be related causally if the latter lies within the light-cone of the earlier.

20.14 FOUR-SPACE ASPECT OF THE LORENTZ TRANSFORMATION

Writing ω for ct, the Lorentz transformation may be expressed as a matrix equation:

$$\begin{bmatrix} x' \\ y' \\ z' \\ \omega' \end{bmatrix} = \begin{bmatrix} \beta & 0 & 0 & \dfrac{j\beta \mathrm{u}}{\mathrm{c}} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\dfrac{j\beta \mathrm{u}}{\mathrm{c}} & 0 & 0 & \beta \end{bmatrix} \begin{bmatrix} x \\ y \\ z_2 \\ \omega \end{bmatrix} \tag{20.67}$$

This linear transformation closely resembles that giving the change to a new set of axes in three-space, i.e. if with respect to $Ox'y'z'$, Ox has the direction cosines (l_1, m_1, n_1), Oy has the direction cosines (l_2, m_2, n_2), and Oz has the direction cosines (l_3, m_3, n_3), then the point (x, y, z) becomes (x', y', z') which is given by the equation:

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} l_1 & m_1 & n_1 \\ l_2 & m_2 & n_2 \\ l_2 & m_3 & n_3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{20.68}$$

In this equation,

$$l_1^2 + m_1^2 + n_1^2 = l_2^2 + m_2^2 + n_2^2 = l_3^2 + m_3^2 + n_3^2 = 1 \tag{20.68a}$$

and

$$l_2l_3 + m_2m_3 + n_2n_3 = l_3l_1 + m_3m_1 + n_3n_1 = l_1l_2 + m_1m_2 + n_1n_2$$

$$= m_1n_1 + m_2n_2 + m_3n_3 = n_1l_1 + n_2l_2 + n_3l_3 = l_1m_1 + l_2m_2 + l_3m_3 = 0 \tag{20.68b}$$

This means that, with respect to $Oxyz$, Ox' has the direct cosines (l_1, l_2, l_3), Oy' has the direction cosines (m_1, m_2, m_3), and Oz' has the direction cosines (n_1, n_2, n_3). This is because $Oxyz$ are mutually perpendicular. Equations (20.67) do not have the corresponding properties; but by a simple alteration they can be made to have such a property. This is by taking the time coordinate as 'imaginary', so that these four are then:

$$x_1 = x, \quad x_2 = y, \quad x_3 = z, \quad x_4 = jct \tag{20.69}$$

This is rather inconvenient for visualizing, because the light-cone is now imaginary; but the Lorentz transformation now becomes:

$$\begin{bmatrix} x_1' \\ x_2' \\ x_3' \\ x_4' \end{bmatrix} = \begin{bmatrix} \beta & 0 & 0 & \dfrac{j\beta \mathrm{u}}{\mathrm{c}} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\dfrac{j\beta \mathrm{u}}{\mathrm{c}} & 0 & 0 & \beta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \tag{20.70}$$

The sum of the squares of the terms in each row and column is now 1; and the sum of the products in any pair of rows and columns is zero. Hence the Lorentz transformation is simply a rotation of the axes in the four-dimensional space described by (x_1, x_2, x_3, x_4).

20.15 FOUR-DIMENSIONAL VECTOR ALGEBRA

A vector having the components A_1, A_2, A_3, A_4 along the four mutually perpendicular axes is called a *4-vector* or *world-vector*. x_1, x_2, x_3, x_4 are themselves the components of a 4-vector—the radius 4-vector; the rules of addition and transformation for any other 4-vector are the same as for x_1, x_2, x_3, x_4. If $\mathbf{i}_1$, $\mathbf{i}_2$, $\mathbf{i}_3$, $\mathbf{i}_4$ are the unit vectors in the 4 directions, then a world-vector $\mathbf{A}$ is given by:

$$\mathbf{A} = \mathbf{i}_1 A_1 + \mathbf{i}_2 A_2 + \mathbf{i}_3 A_3 + \mathbf{i}_4 A_4 \tag{20.71}$$

and, since $\mathbf{i}_1 \cdot \mathbf{i}_1 = \mathbf{i}_2 \cdot \mathbf{i}_2 = \mathbf{i}_3 \cdot \mathbf{i}_3 = \mathbf{i}_4 \cdot \mathbf{i}_4 = 1$, the scalar product of the 4-vectors would be:

$$\mathbf{A} \cdot \mathbf{B} = A_1 B_1 + A_2 B_2 + A_3 B_3 + A_4 B_4 \tag{20.72}$$

It should be noted that $\mathbf{i}_1, \mathbf{i}_2, \mathbf{i}_3, \mathbf{i}_4$ are ordinary vectors, and the scalar product of any two of them, only involves two dimensions. Just as 3 mutually perpendicular vectors add up to a vector in 3-space, so the adding up of four mutually perpendicular vectors give a 'vector' in 4-space or 4-vector.

It should be further noted that 'vector products' cannot be formed in 4-space; if $(\mathbf{i}_1 \times \mathbf{i}_2)$ is perpendicular to $\mathbf{i}_1$ and $\mathbf{i}_2$, it could be anywhere in the plane $(\mathbf{i}_3, \mathbf{i}_4)$.

If, on the analogy of the vector product, we formed the products of the form $(A_i B_j - A_j B_i)$, we should get *six* components, and hence the analogue of $(\mathbf{A} \times \mathbf{B})$ (of 3-dimensions) will not be a 4-vector in 4-dimensions. Thus the mode of generalizing from 3- to 4-dimensions is found by writing

$$A_i B_j - A_j B_i = C_{ij} = -C_{ji} \tag{20.73}$$

and representing the vector product by the anti-symmetric matrix:

$$\begin{bmatrix} 0 & C_{12} & C_{13} \\ C_{21} & 0 & C_{23} \\ C_{31} & C_{32} & 0 \end{bmatrix}$$

By analogy, we can then say:

$$\mathbf{A} \times \mathbf{B} = \begin{bmatrix} 0 & C_{12} & C_{13} & C_{14} \\ C_{21} & 0 & C_{23} & C_{24} \\ C_{31} & C_{32} & 0 & C_{34} \\ C_{41} & C_{42} & C_{43} & 0 \end{bmatrix} \tag{20.74}$$

By analogy with the vector operator del ∇, we introduce the corresponding 4-vector operator:

$$\square \equiv \mathbf{i}_1 \left(\frac{\partial}{\partial x_1} \right) + \mathbf{i}_2 \left(\frac{\partial}{\partial x_2} \right) + \mathbf{i}_3 \left(\frac{\partial}{\partial x_3} \right) + \mathbf{i}_4 \left(\frac{\partial}{\partial x_4} \right) \tag{20.75}$$

We then define:

$$\text{grad } \phi = \square \phi = \mathbf{i}_1 \left(\frac{\partial \phi}{\partial x_1} \right) + \mathbf{i}_2 \left(\frac{\partial \phi}{\partial x_2} \right) + \mathbf{i}_3 \left(\frac{\partial \phi}{\partial x_3} \right) + \mathbf{i}_4 \left(\frac{\partial \phi}{\partial x_4} \right) \tag{20.76}$$

$$\text{div } \mathbf{A} = \square \cdot \mathbf{A} = \frac{\partial A_1}{\partial x_1} + \frac{\partial A_2}{\partial x_2} + \frac{\partial A_3}{\partial x_3} + \frac{\partial A_4}{\partial x_4} \tag{20.77}$$

However, if curl $\mathbf{A} = \square \times \mathbf{A}$, it is not a 4-vector, but an anti-symmetric matrix:

$$\text{curl } \mathbf{A} = \square \times \mathbf{A} = \begin{bmatrix} 0 & F_{12} & F_{13} & F_{14} \\ F_{21} & 0 & F_{23} & F_{24} \\ F_{31} & F_{32} & 0 & F_{34} \\ F_{41} & F_{42} & F_{43} & 0 \end{bmatrix} = [F_{ij}] \tag{20.78}$$

where
$$F_{12} = \frac{\partial A_2}{\partial x_1} - \frac{\partial A_1}{\partial x_2}, \ \ldots, \text{ and so on} \tag{20.79}$$

In 3-dimensions, the components of curl **A** are of the above form and satisfy the identity div curl **A** = 0. Corresponding to this in 4-dimensions, there are four identities which may be expressed in the general form,

$$\frac{\partial F_{jk}}{\partial x_i} + \frac{\partial F_{ki}}{\partial x_j} + \frac{\partial F_{ij}}{\partial x_k} = 0 \tag{20.80}$$

where i, j, k, are any 3 of the 4 numbers 1, 2, 3, 4.

A matrix whose components transform in a certain way when the axes are rotated is called a 'tensor'. $[F_{ij}]$ is one of such tensors, it is a 'four-tensor of the second rank'.

20.16 FOUR-DIMENSIONAL ELECTROMAGNETISM

We have seen that the charge density ρ_C and the components of the current density **J** transformed in the Lorentz transformation in the same way as t, x, y, z. Hence it follows that the quantity defined by

$$\mathbf{J} = \mathbf{i}_1 J_x + \mathbf{i}_2 J_y + \mathbf{i}_3 J_z + \mathbf{i}_4 jc\rho_C \tag{20.81}$$

is a 4-vector. It is called the 'current 4-vector'. In the same way, the three components of the magnetic vector potential **A** combine with the electric scalar potential ϕ to form the *4-potential*:

$$\mathbf{A} = \mathbf{i}_1 A_x + \mathbf{i}_2 A_y + \mathbf{i}_3 A_z + \mathbf{i}_4 \left(\frac{j}{c}\right)\phi \tag{20.82}$$

In both the cases, the dimensional uniformity is preserved by inserting c in the expressions.

We have $\square \times \mathbf{A} = [F_{ij}]$ as defined in Eqs. (20.78) and (20.79).

If, now, we examine the components of $[F_{ij}]$, we find

$$F_{23} = \frac{\partial A_3}{\partial x_2} - \frac{\partial A_2}{\partial x_3} = \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} = B_x \tag{20.83}$$

and

$$F_{41} = \frac{\partial A_4}{\partial x_1} - \frac{\partial A_1}{\partial x_4} = -\left(\frac{j}{c}\right)\left(\frac{\partial A_x}{\partial t}\right) - \left(\frac{j}{c}\right)\left(\frac{\partial \phi}{\partial x}\right)$$

$$= \left(\frac{j}{c}\right) E_x \tag{20.84}$$

using respectively curl $\mathbf{A} = \mathbf{B}$ and $-\text{grad}\ \phi - \left(\frac{\partial A}{\partial t}\right) = \mathbf{E}$.

Thus all the components of $[F_{ij}]$ are proportional to the components of the electromagnetic field; in fact:

$$[F_{ij}] = \begin{bmatrix} 0 & B_3 & -B_3 & -\frac{jE_1}{\mathrm{c}} \\ -B_3 & 0 & B_1 & -\frac{jE_2}{\mathrm{c}} \\ B_2 & -B_1 & 0 & -\frac{jE_3}{\mathrm{c}} \\ \frac{jE_1}{\mathrm{c}} & \frac{jE_2}{\mathrm{c}} & \frac{jE_3}{\mathrm{c}} & 0 \end{bmatrix} \tag{20.85}$$

The components of **E** and **B** are thus all contained in this 'electromagnetic field tensor'.

Next, we have a look at the identities of Eq. (20.80), and substituting $i, j, k = 1, 2, 3$, respectively, we get

$$\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z} = 0 \quad \text{i.e. div } \mathbf{B} = 0 \tag{20.86}$$

and with $i, j, k = 2, 3, 4$, respectively, we get

$$-\left(\frac{j}{\mathrm{c}}\right)\left(\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z}\right) - \left(\frac{j}{\mathrm{c}}\right)\left(\frac{\partial B_x}{\partial t}\right) = 0$$

which is the x-component of the equation:

$$\text{curl } \mathbf{E} = -\left(\frac{\partial \mathbf{B}}{\partial t}\right) \tag{20.87}$$

Thus, these two among the four Maxwell's equations are implied when the field components are expressed in the form:

$$[F_{ij}] = \square \times \mathbf{A} \tag{20.88}$$

This is quite natural and logical; because they are also implied by the underlying equations:

$$\mathbf{E} = -\text{grad}\ \phi - \left(\frac{\partial \mathbf{A}}{\partial t}\right) \quad \text{and} \quad \mathbf{B} = \text{curl } \mathbf{A}$$

The development of the electromagnetic theory in free space is simplified, if **A** is made to satisfy the condition:

$$\text{div } \mathbf{A} + \left(\frac{1}{\mathrm{c}^2}\right)\left(\frac{\partial \phi}{\partial t}\right) = 0 \qquad \text{(i.e. Lorentz gauge)} \tag{20.89}$$

It is easily verified that in 4-space, this is equivalent to:

$$\square \cdot \mathbf{A} = 0 \tag{20.90}$$

The other two Maxwell's equations,

$$\text{div } \mathbf{D} = \rho_{\mathrm{C}} \tag{20.91}$$

and

$$\text{curl } \mathbf{H} - \left(\frac{\partial \mathbf{D}}{\partial t}\right) = \mathbf{J} \tag{20.92}$$

contain **D**, **H**, and the components of the 4-current. So writing in the Cartesian form with numbered components, we get

$$\left(\frac{\partial H_3}{\partial x_2}\right) - \left(\frac{\partial H_2}{\partial x_3}\right) - (jc)\left(\frac{\partial D_1}{\partial x_4}\right) = J_1 \tag{20.93a}$$

$$\left(\frac{\partial H_1}{\partial x_3}\right) - \left(\frac{\partial H_3}{\partial x_1}\right) - (jc)\left(\frac{\partial D_2}{\partial x_4}\right) = J_2 \tag{20.93b}$$

$$\left(\frac{\partial H_2}{\partial x_1}\right) - \left(\frac{\partial H_1}{\partial x_2}\right) - (jc)\left(\frac{\partial D_3}{\partial x_4}\right) = J_3 \tag{20.93c}$$

$$(jc)\left(\frac{\partial D_1}{\partial x_1} + \frac{\partial D_2}{\partial x_2} + \frac{\partial D_3}{\partial x_3}\right) = J_4 (= jc\rho_C) \tag{20.93d}$$

If, now, we define a second field tensor

$$[G_{ij}] = \begin{bmatrix} 0 & H_3 & -H_2 & -jcD_1 \\ -H_3 & 0 & H_1 & -jcD_2 \\ H_2 & -H_1 & 0 & -jcD_3 \\ jcD_1 & jcD_2 & jcD_3 & 0 \end{bmatrix} \tag{20.94}$$

then Eqs. (20.93) reduce to the form:

$$\frac{\partial G_{11}}{\partial x_1} + \frac{\partial G_{12}}{\partial x_2} + \frac{\partial G_{13}}{\partial x_3} + \frac{\partial G_{14}}{\partial x_4} = J_1, \text{ and so on.}$$

Or, in a more compact form:

$$\sum_{j=1,2,\cdots}^{4} \left(\frac{\partial G_{ij}}{\partial x_j}\right) = J_i \quad (i = 1, 2, 3, 4) \tag{20.95}$$

The 4-space approach thus enables the four field vectors **E**, **B**, **D,** and **H** to be replaced by two field tensors $[F_{ij}]$ and $[G_{ij}]$; the four Maxwell's equations to be replaced by two sets of equations (20.80) or (20.78) and (20.95), the vector and the scalar potentials by a single 4-potential; the two equations $\mathbf{E} = -\text{ grad}\,\phi - (\partial \mathbf{A}/\partial t)$ and $\mathbf{B} = \text{curl } \mathbf{A}$ by the single Eq. (20.88). Finally, the continuity equation

$$\text{div } \mathbf{J} + \frac{\partial \rho_C}{\partial t} = 0 \tag{20.96}$$

is implicit in Eqs. (20.91) and (20.92), and these take the single form

$$\square \cdot \mathbf{J} = 0 \tag{20.97}$$

20.17 INVARIANCE OF MAXWELL'S EQUATIONS UNDER THE LORENTZ TRANSFORMATION

From the Maxwell's equations, together with the relations $\mathbf{D} = \varepsilon_0\mathbf{E}$ and $\mathbf{B} = \mu_0\mathbf{H}$, we conclude that the electromagnetic waves travel in free space with velocity c. The Lorentz transformation assures the truth of this statement for all inertial frames of reference. It is therefore to be expected that Maxwell's equations will hold good in all inertial frames, i.e. they will be invariant under a Lorentz transformation.

A direct proof of this statement is possible, which is rather laborious, by writing the equations in Cartesian coordinates, and then transforming them to the 'primed' system. A more fundamental approach starts from the 4-dimensional field tensors. Comparison of Eqs. (20.85) and (20.94) shows that, in free space, the terms of the tensors $[F_{ij}]$ and $[G_{ij}]$ are proportional:

$$F_{ij} = \mu_0 G_{ij} \tag{20.98}$$

Thus, Eq. (20.95) may be rewritten as

$$\sum_{j=1,2,\cdots}^{4} \left(\frac{\partial F_{ij}}{\partial x_j} \right) = \mu_0 J_i \quad (i = 1, 2, 3, 4) \tag{20.99}$$

Substituting,

$$F_{ij} = \frac{\partial A_j}{\partial x_i} - \frac{\partial A_i}{\partial x_j}, \qquad \text{from Eq. (20.79), we get}$$

$$\sum_{j=1,2,\cdots}^{4} \left(\frac{\partial^2 A_j}{\partial x_j \partial x_i} \right) - \sum_{j=1,2,\cdots}^{4} \left(\frac{\partial^2 A_i}{\partial x_j^2} \right) = \mu_0 J_i \tag{20.100}$$

and the first term is

$$\left(\frac{\partial}{\partial x_i} \right) (\square \cdot \mathbf{A}) = 0$$

when we assume that

$$\text{div } \mathbf{A} + \left(\frac{1}{c^2} \right) \left(\frac{\partial \phi}{\partial t} \right) = 0 \quad \text{[see Eqs. (20.89) and (20.90)]}$$

Thus Eq. (20.100) reduces to the four equations,

$$\sum_{j=1,2,\cdots}^{4} \left(\frac{\partial^2 A_i}{\partial x_j^2} \right) = -\mu_0 J_i$$

or

$$\square \cdot \square A_i = -\mu_0 J_i \quad (i = 1, 2, 3, 4) \tag{20.101}$$

Here A_i is treated as a set of 4 scalars. If a Lorentz transformation is performed, A_i and J_i transform in the same way, and the transformed components obey the same equation.

If we substitute the components A_i and J_i from Eqs. (20.81) and (20.82), we find that the first three of Eqs. (20.101) are

$$\nabla^2 \mathbf{A} - \left(\frac{1}{c^2} \right) \left(\frac{\partial^2 \mathbf{A}}{\partial t^2} \right) = -\mu_0 \mathbf{J} \tag{20.102a}$$

and the fourth

$$\nabla^2\phi - \left(\frac{1}{c^2}\right)\left(\frac{\partial^2\phi}{\partial t^2}\right) = -\frac{\rho_C}{\varepsilon_0} \tag{20.102b}$$

These are the equations which give rise to the waves propagating with the velocity c; their similarity, surprising in a 3-dimensional approach, is here quite natural.

Thus the Maxwell's equations in free space have led to a set of invariant equations (20.101), which lead in turn to waves that propagated with velocity c. Maxwell's equations transcend the Lorentz transformation, and are so built deeply into the world structure. They are, more fundamental, for example, than the Newton's second law of motion, which is usually stated in a form which assumes constancy of mass.

20.18 RELATIVISTIC ASPECTS OF ELECTROMAGNETIC INDUCTION

We shall conclude our discussion of the relativistic approach to electromagnetic field theory, by considering the practical problem of 'electromagnetic induction'.

When induction occurs in a magnetic field of constant pattern but varying intensity (as in a transformer), we have the so called *transformer induction* which is a sign of the fact that every varying magnetic field is accompanied by an electric field (also varying). This is expressed in the Maxwell's equation:

$$\text{curl } \mathbf{E} = -\left(\frac{\partial \mathbf{B}}{\partial t}\right)$$

When we have a field pattern, which moves but remains constant in intensity (e.g. that set up by rotating magnetic poles in an alternator), we have the *motional induction*, and find it convenient to apply the *flux-cutting method*. This is a relativistic phenomenon. In Figure 20.11, in the frame F' (which is attached to the rotating pole), we have only a magnetic field; but in the frame F (attached to the stator), we have also an electric field, which makes itself known as an emf induced in a winding in that frame F.

The emf can equally well be calculated by the flux-linking approach [$\mathscr{E} = -(d\Lambda/dt)$], which is equivalent to the flux-cutting ($\mathscr{E} = Blu$). But the equivalence seems to break down when we have *homopolar generator*, with no passage from N (north pole) to S (south pole) poles.

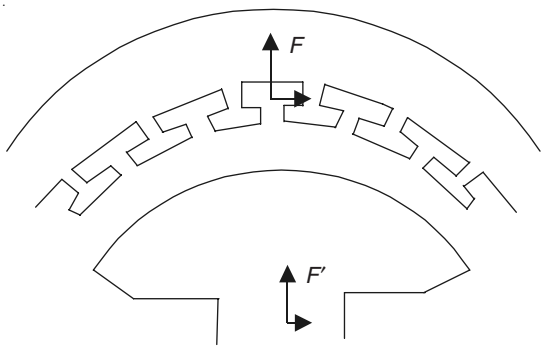


Figure 20.11 Stator and rotor of a rotating alternator.

20.18.1 Electromagnetic Induction by Flux-cutting

This phenomenon takes a new aspect when viewed relativistically. For instance, the train of arguments describing the operation of a dynamo with its field (winding) on the stator becomes:

1. The field winding sets up B with respect to the stator (frame F).
2. Both E' and B' are potentially present in the rotor (frame F').
3. E' is short-circuited by the rotor conductors, and is reduced to zero by the movement of the charges to the ends of the conductors, just as when a metal rod is placed in an electrostatic field.
4. The reduction of E' to zero brings into being an electric field E with respect to the stator (the frame F). The sources of the field are the charges aforementioned. Its magnitude is given by:

$$\mathbf{E} = -(\mathbf{u} \times \mathbf{B}') = -(\mathbf{u} \times \mathbf{B}), \qquad \text{to first order} \tag{20.103}$$

This gives with respect to (the frame) F, a potential difference between the ends of the conductor; but $\mathbf{E} \cdot d\mathbf{l} = 0$ for a fixed circuit with no change of flux-linkage, and so we get an equal and opposite contribution across a measuring instrument. The flux-cutting action is shown diagrammatically in Figure 20.12.

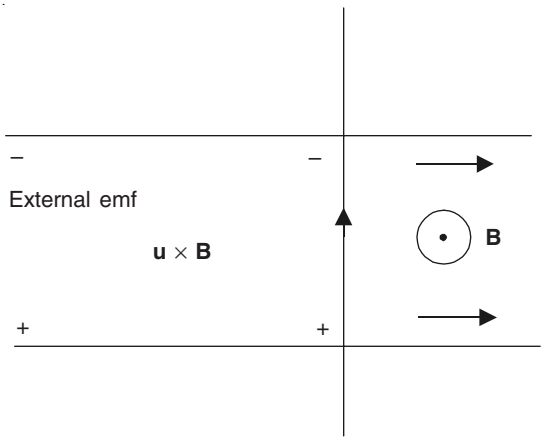


Figure 20.12 Flux-cutting emf.

20.18.2 The Concept of a Moving Field

There is no way in which the ideas of 'fixity' or 'motion' can be ascribed to a field. The velocity **u** in the Lorentz formula is not 'velocity relative to the field' but 'velocity relative to the observer'; for another observer with relative motion, **E**, **B**, and **u** would be different, yet in such a way as to make **F** (the Lorentz force) the same. What, then, are we doing when we make use of the idea of a moving or gliding field, for example in an ac machine?

We now consider the induction of emf in the stator coil of an alternator (Figure 20.13). We could calculate it by flux-linkage, but in fact we usually use flux cutting. regarding the flux as glued to the magnetic pole. This is equivalent to changing from a frame F which is fixed with respect to the stator, in which the flux density at any point is changing, to a frame F', fixed with respect to the rotor, in which the flux pattern is practically unchanging. Relative to the frame F', the stator conductors are moving through an unchanging field—a fair case for a flux-cutting method.

We are not necessarily ascribing to the field the same motion as that of the coils which produce it. Thus the words 'the velocity of the field' must be taken to mean 'the velocity of a frame of reference, with respect to which the field pattern is unchanging'. In an induction motor, the frame F' (with respect to which the field is unchanging) glides round at synchronous speed, which is not the speed of any material portion of the motor. When the field is pulsating, whatever the frame of reference we may choose, the 'moving field' idea has reached its limit of usefulness.

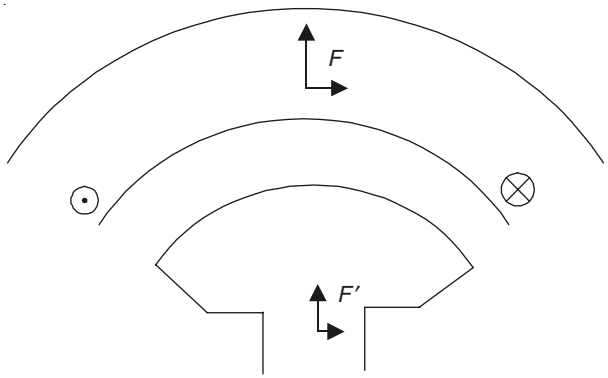


Figure 20.13 Stator and rotor of an alternator to explain the concept of a moving field.

20.18.3 Examples of Homopolar Induction

20.18.3.1 Faraday's disc

When a disc is rotated in an axial magnetic field, an emf is observed between the periphery of the disc and the axle [Figure 20.14(a)]. The value of this induced emf is the value given by 'flux cutting' by any radius of the disc; the method of 'flux linking' appears to give zero emf.

Take the axes as shown in Figure 20.14(b), and let the two frames F and F' momentarily coincide. The frame F is fixed in space, and the frame F' is fixed to the disc. The relative velocity $u = \omega r$ is in the x-direction.

As the disc is a conductor, there will be no electric field in the x- or y-direction relative to the frame F' (which is fixed to the disc).

$$\therefore \qquad E'_x = E'_y = 0 \tag{20.104}$$

But
$$E'_x = E_x \qquad \text{and} \qquad E'_y = E_y - uB_z \tag{20.105}$$

where B_z is the axial, imposed magnetic field.

Hence, with respect to the frame F (which is fixed in space),

$$E_x = 0 \qquad \text{and} \qquad E_y = uB_z \tag{20.106}$$

$\therefore$ The emf across the radius PQ [Figure 20.14(a)] is

$$= \int_{r=0}^{r=a} \omega r B_z dr = \left(\frac{1}{2}\right)\omega a^2 B_z \tag{20.107}$$

where a is the radius of the disc. This emf value is measured by the meter.

The *amendment* to the 'flux-cutting' rule which will make it cover this and similar cases is that the circuit must be such that at no place are the particles of the material moving across it. Such a circuit is shown in Figure 20.14(c), and it gives the correct answer when the flux-linking rule is applied.

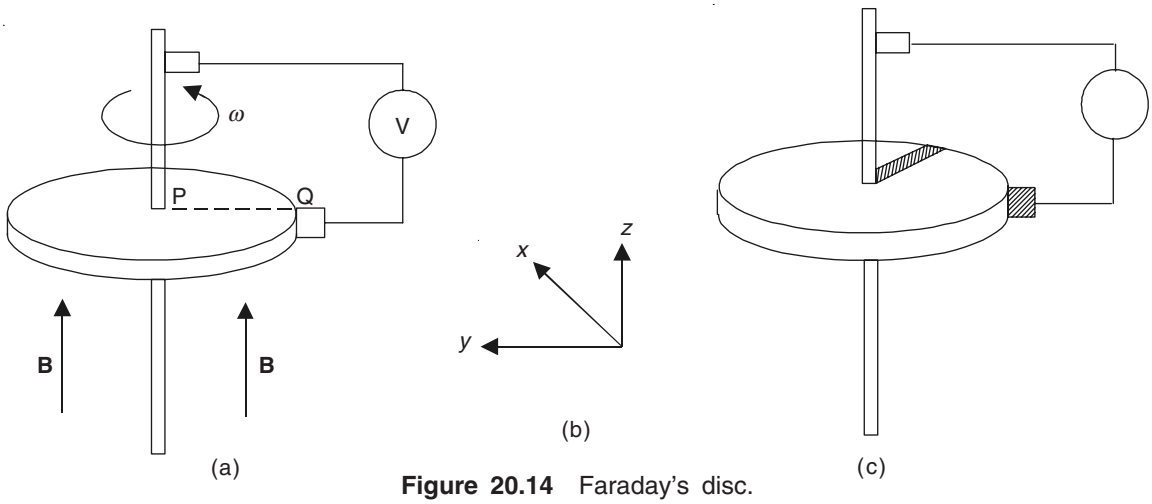


Figure 20.14 Faraday's disc.

20.18.3.2 Faraday's disc with rotating magnet

This is a slight modification of the previous configuration, as shown in Figure 20.15. The question now is: 'if the Faraday's disc is excited by a cylindrical bar magnet, which rotates with it, what result is expected?'

Actually, the induced emf is unaffected by the rotation of the magnet. This is because the term 'rotating field' is fundamentally meaningless when the field pattern is symmetrical.

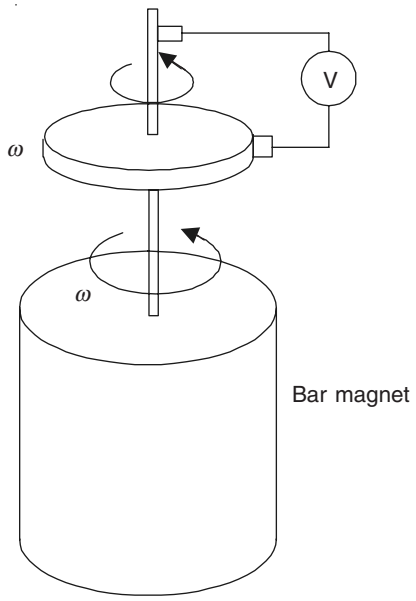


Figure 20.15 Faraday's disc with rotating cylindrical bar magnet.

20.18.3.3 Induction in a rotating cylindrical bar magnet

This problem is akin to the last one. Here, do we expect an induced emf? (Figure 20.16)

The answer to this question is, yes. Since a symmetrical rotating magnetic field is meaningless, we should get the same answer whether we assume it or not.

If the field is stationary, the induction occurs across the Faraday disc. If the field rotates, we get an induced emf in the external circuit. It can be seen that the two mental pictures give emf of the same polarity as also of the same magnitude.

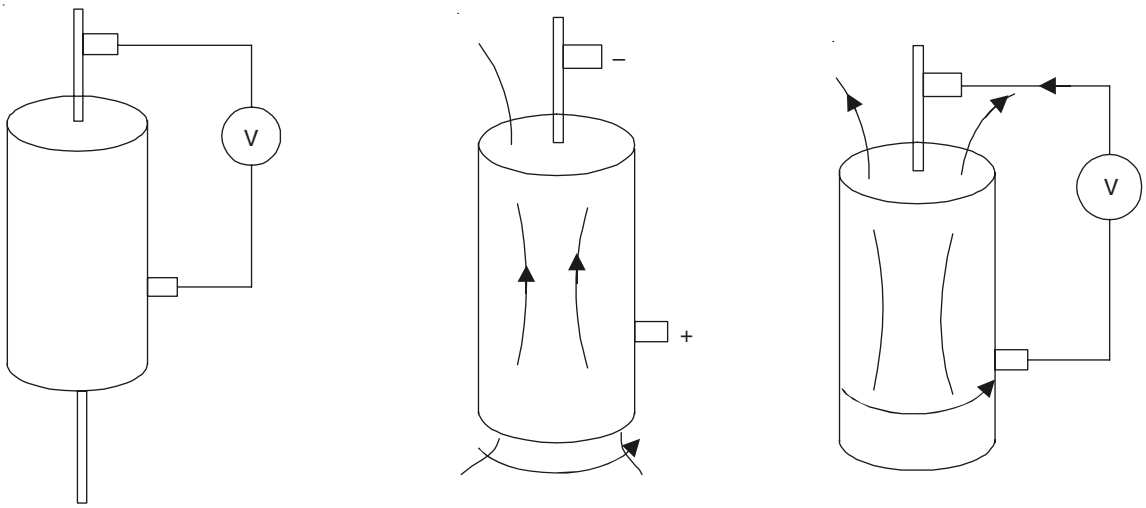


Figure 20.16 A rotating bar magnet.

20.18.3.4 Current, terminal voltage, and power in a Faraday disc generator

The Faraday disc generator (or Homopolar or acyclic generator), with its details, is shown in Figure 20.17. The applied flux density B_0 is constant, and the shaft is driven at constant velocity (angular) ω. The material of the disc is homogeneous, isotropic and electrically linear with material constants σ, μ_0, ε_0. The dimensions are defined in Figure 20.17. We shall derive the expressions for the terminal voltage and the current for any load resistance R_L (under steady-state operation).

It should be noted that the current in the disc is radial, and the current density is uniform around the periphery at any radius. The magnetic field generated by the current density is tangential and has no effect on the terminal voltage. Hence the field due to the current in the disc is neglected.

In the cylindrical coordinate system chosen, the symmetry and the uniformity of the problem shows that we can assume

$$\frac{\partial}{\partial \phi} = \frac{\partial}{\partial z} = 0$$

and the only variation is along the radius (= r).

If the total output current is I, the radial component of the current density in the disc (at a radial distance r) is given by

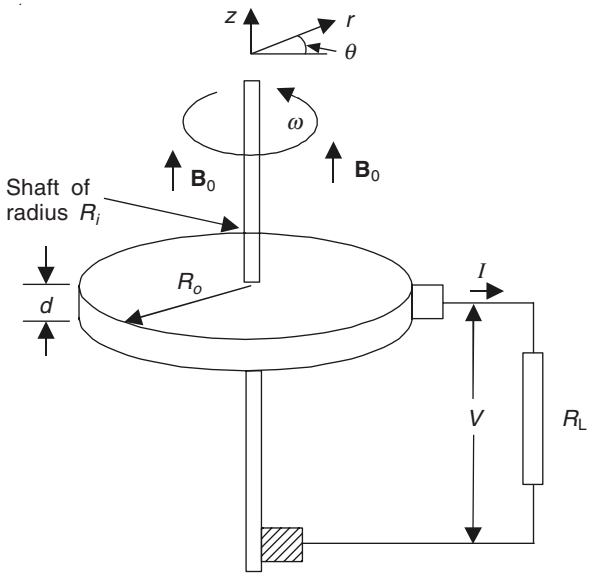


Figure 20.17 Faraday's disc generator, connected to a load resistor.

$$J_r = \frac{I}{2\pi r d} \tag{20.108}$$

By Ohm's law for a grain of matter at a radius r, we get the relationship

$$J_r = \sigma (E_r + \omega r B_0) \tag{20.109}$$

where B_0 is the magnitude of the externally applied magnetic field, and E_r the radial component of the electric field intensity.

The tangential component of **B** (the flux density B_ϕ due to the current flow in the disc) is parallel to the material velocity, and hence does not contribute to the ($\mathbf{v} \times \mathbf{B}$) term above. Thus the neglect of this component of **B** generated by the current flow in the disc is justified.

$$\therefore \qquad E_r = \frac{I}{2\pi r d \sigma} - \omega r B_0 \tag{20.110}$$

Since there is no time-rate of change of **B** in the fixed reference frame, the terminal voltage is

$$V = -\int_{r=R_i}^{r=R_o} E_r dr = -\left(\frac{I}{2\pi\sigma d}\right)\ln\left(\frac{R_o}{R_i}\right) + \left(\frac{\omega B_0}{2}\right)\left(R_o^2 - R_i^2\right) \tag{20.111}$$

In the above expression, if we let $R_L \to \infty$, then the open-circuit voltage is

$$V_{oc} = \left(\frac{\omega B_0}{2}\right)\left(R_o^2 - R_i^2\right) \tag{20.112}$$

and the internal resistance of the generator is

$$R_{int} = \frac{\ln\left(\frac{R_o}{R_i}\right)}{2\pi\sigma d} \tag{20.113}$$

Hence the terminal relationship is

$$I = \frac{V_{\text{oc}}}{R_{\text{L}} + R_{\text{int}}} \tag{20.114}$$

The short-circuit current is

$$I_{\text{sc}} = \frac{V_{\text{oc}}}{R_{\text{int}}} = \frac{\omega B_0 (R_o^2 - R_i^2) 2\pi\sigma d}{2 \ln\left(\frac{R_o}{R_i}\right)} \tag{20.115}$$

The maximum power delivered by the generator is

$$P_{\text{max}} = \frac{V_{\text{oc}} I_{\text{sc}}}{4}$$

$$= \frac{\omega^2 B_0^2 (R_o^2 - R_i^2) \pi\sigma d}{8 \ln\left(\frac{R_o}{R_i}\right)} \tag{20.116}$$

20.18.4 Herring's Experiment

This experiment is one in which the terminals of a galvanometer are connected to a pair of springy metal strips, as shown in Figure 20.18, which are held apart at one end by an insulating block. But the strips are so shaped that their springiness makes them snap together (bringing them in contact) at the free end, thus short-circuiting the galvanometer.

In the experiment, the free ends of the strips are forced apart by pushing a permanent magnet (say, of cylindrical cross-section) between them side-ways; when the magnet has been completely pushed in, the contacts spring together again, with the magnetic flux from the magnet now linking the circuit.

The question is: do we expect an induced emf, when the cylindrical bar magnet is pushed between the clips?

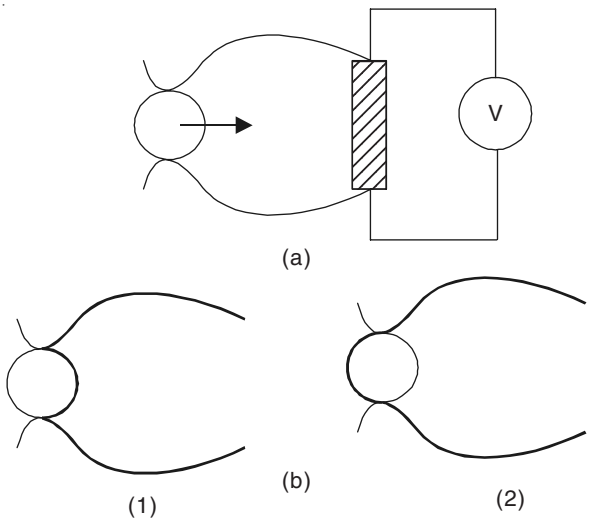


Figure 20.18 Herring's experiment.

The answer is: there is no reading in the galvanometer. The reason is that. as in the previous case of the Faraday's disc, the flux-linkage must be calculated in a circuit *across which the material particles are not moving.* Two such circuits are shown in the Figure 20.18(b); and in neither is there any change in the flux-linkage.

It will be found that when the idea of a *moving field* is successfully employed. we are really talking about a *moving field pattern*. This means transference from a frame of reference in which the field is unchanging to another in which the field-pattern moves without variation of shape or magnitude. If no frame exists in which the field is unchanging, the concept of a moving field is a snare.

20.18.5 The Cullwick Experiment

A piece of steel tubing R (as shown in Figure 20.19) is coaxial with a long wire C carrying a direct current *I*. The tubing moves with a uniform velocity u in the direction opposite to that of the current flow.

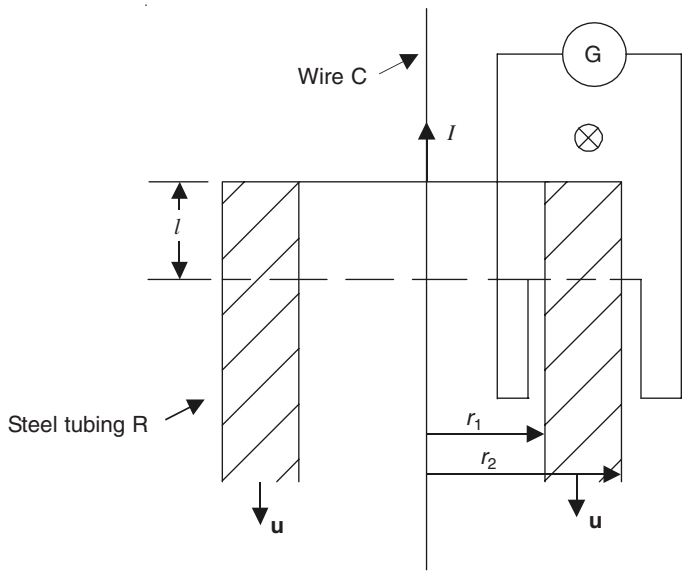


Figure 20.19 The Cullwick experiment.

The questions in this experiment are:

1. What is the induced emf (in the loop or the contour shown in Figure 20.19)?
2. Does the emf change, if the steel tubing is replaced by a brass tubing (of identical shape and size)?

A. Direct solution. We consider the emf in the loop shown in Figure 20.19, comprising a path round the edge of the cylinder.

In air, $$B = \frac{\mu_0 I}{2\pi r} \tag{20.117}$$

If the cylinder has fallen through a height l in the time T,

$$\text{the extra flux in the length } l = \left(\frac{\mu_0 Il}{2\pi}\right)\ln\left(\frac{r_2}{r_1}\right) \tag{20.117a}$$

Since this fall has taken place in the time T,

$$\text{the induced emf} = \left(\frac{\mu_0 Il}{2\pi T}\right)\ln\left(\frac{r_2}{r_1}\right) \tag{20.118}$$

and

$$\text{the current in the loop} = \left(\frac{\mu_0 Il}{2\pi RT}\right)\ln\left(\frac{r_2}{r_1}\right) \tag{20.119}$$

where R is the resistance of the loop.

And

$$\text{the charge} = \left(\frac{\mu_0 Il}{2\pi R}\right)\ln\left(\frac{r_2}{r_1}\right) \tag{20.120}$$

This is *independent of the material* of the cylinder.

B. Moving frame of reference solution. (Figure 20.20)

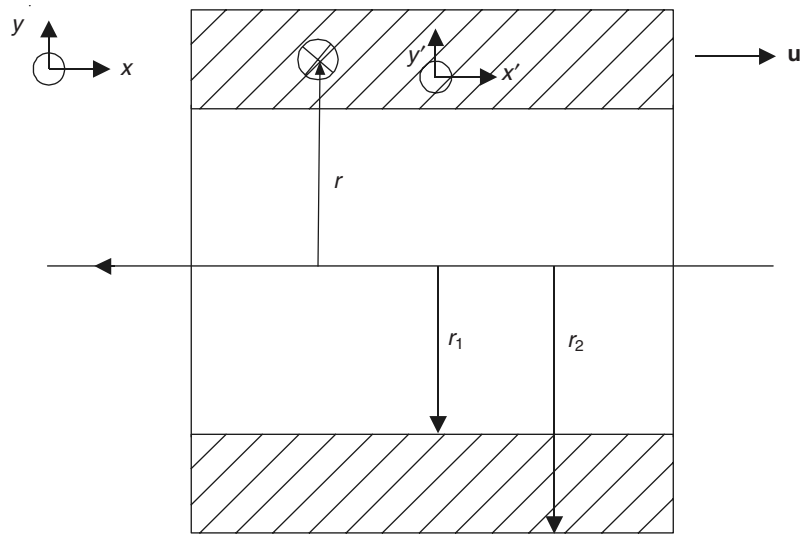


Figure 20.20 The Cullwick experiment, showing the frames of reference.

In the stationary frame F

$$H = H_\phi = \frac{I}{2\pi r} \quad (= -H_z) \tag{20.121}$$

In the moving frame F', there is potentially

$$D' = \left(\frac{1}{c^2}\right)(\mathbf{u} \times \mathbf{H}) \tag{20.122}$$

which gives

$$D'_y = \frac{uI}{2\pi c^2 r} \quad \text{or} \quad E'_y = \left(\frac{\mu_0}{2\pi}\right)\left(\frac{uI}{r}\right) \tag{20.123}$$

This is short-circuited by the conductor, giving rise to an emf:

$$\mathscr{E} = \int_{r=r_1}^{r=r_2} E'_y \, dr = \left(\frac{\mu_0 \mathrm{u} I}{2\pi}\right) \ln\left(\frac{r_2}{r_1}\right) \tag{20.124}$$

If the velocity $\mathrm{u} = \dfrac{dx}{dt}$, this gives a charge:

$$= \int_0^T \left(\frac{\mathscr{E}}{R}\right) dt = \left(\frac{\mu_0 I}{2\pi R}\right) \ln\left(\frac{r_2}{r_1}\right) \int_0^T \left(\frac{dx}{dt}\right) dt$$

$$= \left[\left(\frac{\mu_0 I}{2\pi R}\right) \ln\left(\frac{r_2}{r_1}\right)\right] \times \text{distance moved} \tag{20.125}$$

where R is the resistance of the circuit.

This argument is independent of the material of the cylinder.

PROBLEMS

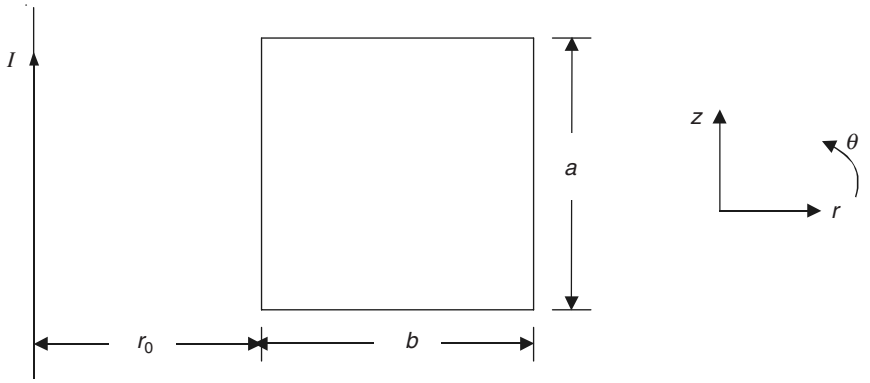
20.1 A car with 2 metre long bumper is travelling at 100 km/h. Find the potential drop produced in the bumper due to earth's magnetic field of 3.2×10^{-5} webers/m^2 and the angle of dip of 64° 9′.

Ans.: 1.6 mV

20.2 The differential form of Faraday's law when applied to a system moving with a velocity **u** can be written as:

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} + \nabla \times \mathbf{u} \times \mathbf{B}$$

A rectangular loop is located in the field of a long current-carrying straight wire as shown below.



Using the above expression, evaluate:

(i) The induced voltage in the loop if it is fixed in space but I varies as $I_0 \cos \omega t$.

(ii) The magnitude and the direction of the induced voltage as a function of r when $I = I_0$ and is constant, but the loop moves towards the wire with a velocity **u**.

(iii) The induced voltage in the loop when it moves towards the long conductor with a velocity **u** and the current varies as $I = I_0 \cos \omega t$.

Ans.: (i) $\dfrac{\mu_0 \omega I_0}{2\pi} a \ln \dfrac{r_0 + b}{r_0} \cdot \sin \omega t$; (ii) $-\dfrac{\mu_0 \mathrm{u} I_0 a}{2\pi}\left(\dfrac{1}{r_0} - \dfrac{1}{r_0 + b}\right)$;

(iii) $\dfrac{\mu_0 I_0 a}{2\pi}\left\{\omega \ln \dfrac{r_0 + b}{r_0} \cdot \sin \omega t - \mu_0 \left(\dfrac{1}{r_0} - \dfrac{1}{r_0 + b}\right)\cos \omega t\right\}$.

20.3 The electric and magnetic fields in a region are $\mathbf{E} = \mathbf{i}_y E_0$ and $\mathbf{B} = \mathbf{i}_z B_0$ respectively, where E_0 and B_0 are constants. A small test charge Q having a mass m starts from rest at the origin at the instant $t = 0$.

Using the Lorentz force equation

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

and the equation of motion, show that the velocity components of the charge will be:

$$\mathrm{v}_x = \frac{E_0}{B_0}(1 - \cos \omega_\mathrm{c} t)$$

$$\mathrm{v}_y = \frac{E_0}{B_0} \sin \omega_\mathrm{c} t$$

where $\omega_\mathrm{c} = Q\dfrac{B_0}{m}$.

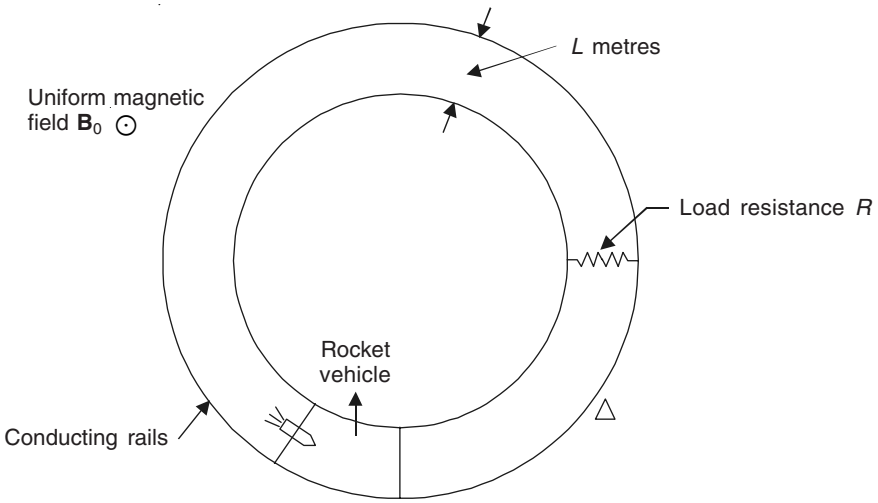
What will be the electric field as seen by an observer moving with the test charge?

20.4 A metal vehicle travels round a set of perfectly conducting rails which form a large circle. The rails are L metres apart and there is a uniform magnetic field B_0 normal to their plane as shown in the figure below. The mass of the vehicle is m, and it is driven by a rocket engine having a constant thrust F_0. The system acts as a dc generator whose output is fed into a load resistance R. Show that the output current I from the system increases exponentially as given by the equation

$$I = \frac{V}{R} = \frac{F_0}{B_0 L}\left[1 - \exp\left\{-t\left(\frac{B_0^2 L^2}{mR}\right)\right\}\right]$$

where V is the induced voltage.

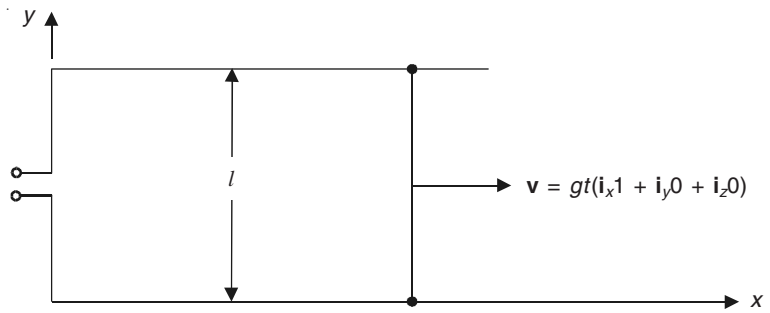
$\left\{Hint: \int \dfrac{dx}{b - ax} = \dfrac{1}{a} \ln (b - ax).\ \text{Use the equation of motion of the cart.}\right\}$

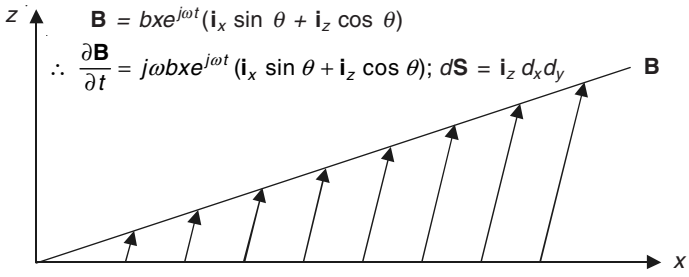


20.5 Faraday's law when applied to a system moving with a velocity v can be written as

$$\mathscr{E} = \oint_C \mathbf{E} \cdot d\mathbf{l} = -\frac{\partial}{\partial t} \iint_S \mathbf{B} \cdot d\mathbf{S} + \oint_C (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{l}$$

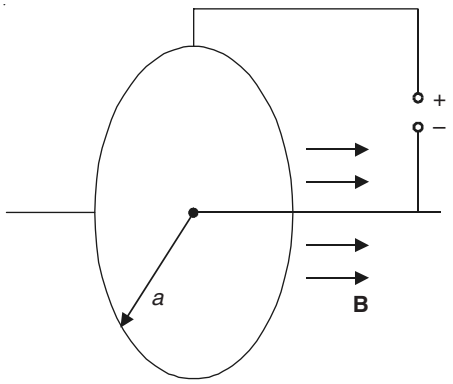
Hence or otherwise, find the emf induced in a rectangular loop, one side of which moves with a constant acceleration g across a flux-density field **B** which increases linearly in the direction of motion of the conductor and periodically with time. The direction of the motion and of the magnetic field are shown in the figures below.





Ans.: $= -e^{j\omega t}\, bl \cos\theta \dfrac{g^2 t^3}{2}\left(1 + \dfrac{j\omega t}{4}\right)$

20.6 The Faraday disc shown is to be used as a motor by including a battery in the circuit. If the current flowing is I and the magnetic flux density **B** is uniform, show that the torque exerted on the disc is $\frac{\phi I}{2\pi}$, where ϕ is the flux crossing the whole disc.



Find I if the battery emf is $\mathscr{E}$, the resistance of the circuit is R, and ω is the angular velocity of the disc in rads/sec.

Ans.: $I = \mathscr{E} - \frac{\phi\omega/2\pi}{R}$

21 Numerical Methods for and Computer Solutions of Electromagnetic Field Problems

21.1 INTRODUCTION

So far in our study of electromagnetic field theory, our main emphasis has been on the proper understanding of the underlying physical concepts of the theory, so that correct mathematical techniques may be applied to solve practical problems. In most of the practical problems, the main difficulty lies not in the selection of the mathematical method, but in the representation of the correct boundary conditions. The interpretation of the physical boundaries, and finding their correct mathematical equivalents, is in fact, the first critical step in solving such problems. This is the reason as to why we have emphasized so strongly on the need for a deep insight into the physics of most electromagnetic phenomena. Having completed the mathematical analysis of the problem, again re-interpreting the mathematical solution in terms of its physics depends very much on the earlier step mentioned above. Hence the correct formulation of the boundary conditions for a specified problem is a very critical step in solving it.

We have discussed earlier in a number of chapters (i.e. Chapters 4, 5, 9), the different methods of solving electrostatic, magnetostatic, and quasi-static magnetic field problems, apart from the various problems (including electrostatic, magnetostatic, and electromagnetic) discussed and solved in different other chapters in this book. It has also been pointed out earlier in this book that whilst the electrostatic and the magnetostatic problems are mostly Laplacian and Poissonian in nature, the electromagnetic problems (i.e. eddy currents and wave problems) are more complex in nature, though under special simplified and idealized conditions, these can be reduced to simpler equations (i.e. Laplace and Poisson). However it should also be remembered that in all these operating differential equations, there is the Laplace's operator ($\equiv \nabla^2$) as the main spatial operator on their left-hand side. Hence the basic methods of solving the general electromagnetic field problems would apply to the static problems, as these can be considered as restricted cases of the more general class.

We have already discussed a number of eddy current and wave problems in the Chapters 15–18, for which we have obtained (mostly) analytical closed-form solutions, using a

number of analytical methods. But these are, by and large, idealized problems. So in this chapter, we shall not go into further details of such methods. This is not to imply that we have covered the complete range of such problems solved by these analytical methods. But since this is an introductory text book on the subject of electromagnetism, we shall not go into the details of those more complex problems, which, of course, would be the subject matter of more specialized texts. Of the other methods which have been discussed earlier, the numerical methods are being increasingly used to solve more and more problems because of the rapid development of more powerful and faster computers. While describing the applications of the numerical methods (i.e. the finite difference method = F.D.M., and the finite element method = F.E.M.), we have pointed out that both these methods are being widely used to solve electrostatic problems. Whilst F.E.M. has gained wide popularity in solving electromagnetic, magnetostatic, and quasi-static magnetic field problems, the F.D.M. has been in significant use for solving electrostatic (and heat transfer) problems, because in electrostatics, there are quite a large range of problems with open boundaries. For such problems, F.D.M. is found to have significant advantages, and the geometry of electrostatic problems (in a large number of cases) permits the use of regular and graded meshes without taking too much computer time. Since we have already discussed in fair detail the methods of F.D.M. (while considering the solutions of electrostatic problems), we do not propose to go into the details of this method for solving magnetic and electromagnetic problems, as the techniques of solving such problems are identical with those for the electrostatic problems. This is true particularly for two-dimensional problems, because even though the vector potential is being used for these problems, the potential usually has only one component (though some of the magnetic field vectors can have more than one component) and hence the treatment is similar to thar of the scalar potential (which is the usual operating variable used in most of the electrostatic problems).

Similar arguments may apply to the use of the F.E.M., though of course, the formulation of the 'functional' would be different for (say) the magnetostatic fields (being a resticted case) from that for the electromagnetic eddy current fields to electromagnetic wave patterns. So we shall now discuss the functionals for various types of electromagnetic field problems, using the concept of magnetic vector potential.

21.2 FINITE ELEMENT METHOD APPLIED TO TWO-DIMENSIONAL ELECTROMAGNETIC FIELD PROBLEMS (INVOLVING EDDY CURRENTS)

As mentioned earlier, F.E.M. is a popular numerical method based on the variational principle for solving any boundary-value field problem. A 'functional' is an integral expression involving a function dependent on the unknown variable and its derivatives. The functional is identified, and minimized with respect to the unknown variables, subject to the boundary conditions specified. This process results in the required solution to the field equations. Let the functional be defined as follows:

$$\Gamma = \iint F(x, y, A, A_x, A_y)\, dx\, dy \tag{21.1}$$

where

$$A_x = \frac{\partial A}{\partial x}, \quad A_y = \frac{\partial A}{\partial y}$$

(*Notes:* 1. It should be carefully noted that here the suffices x and y in A_x, A_y respectively represent the derivatives of A with respect to those independent variables, and *not* the vector components. 2. Also this Γ should not be confused with the reflection coefficient of waveguides, even though the same symbol is used for both.)

As stated before, we are restricting ourselves to two-dimensional problems with the magnetic vector potential having one component only, and the present derivation is in Cartesian coordinate system (though this is not a restriction imposed on the method).

Since the functional is an area integral in terms of the variables x and y within the specified boundaries of the problem, the variations can be effected in A only. Let A be changed by a small quantity δA. The value of the functional then changes to:

$$\Gamma' = \iint F(x, y, A + \delta A, A_x + \delta A_x, A_y + \delta A_y) dx\, dy$$

$\therefore$ The change in the functional is given by:

$$\delta\Gamma = \iint F(x, y, A + \delta A, A_x + \delta A_x, A_y + \delta A_y) dx\, dy - \iint F(x, y, A, A_x, A_y) dx\, dy$$

The first term can be expanded by the Taylor series; and then for sufficiently small values of δA, we can neglect the terms containing higher orders of δA.

$$\therefore \quad \delta\Gamma = \iint \left(\frac{\partial F}{\partial A}\right)\delta A\, dx\, dy + \iint \left(\frac{\partial F}{\partial A_x}\right)\delta A_x\, dx\, dy + \iint \left(\frac{\partial F}{\partial A_y}\right)\delta A_y\, dx\, dy$$

$$= \iint \left(\frac{\partial F}{\partial A}\right)\delta A\, dx\, dy + \iint \left(\frac{\partial F}{\partial A_x}\right) d\left(\frac{\partial A}{\partial x}\right) dx\, dy + \iint \left(\frac{\partial F}{\partial A_y}\right) d\left(\frac{\partial A}{\partial y}\right) dx\, dy \qquad (21.2)$$

From Eq. (21.1),

$$\delta\Gamma = \iint \left(\frac{\partial F}{\partial A}\right)\delta A\, dx\, dy + \iint \left(\frac{\partial F}{\partial A_x}\right)\left(\frac{\partial}{\partial x}\right)(\delta A)\, dx\, dy + \iint \left(\frac{\partial F}{\partial A_y}\right)\left(\frac{\partial}{\partial y}\right)(\delta A)\, dx\, dy$$

But the total differential of δA is given by:

$$d(\delta A) = \left(\frac{\partial}{\partial x}\right)(\delta A)\, dx + \left(\frac{\partial}{\partial y}\right)(\delta A)\, dy \qquad (21.3)$$

$$\therefore \quad \delta\Gamma = \iint \left(\frac{\partial F}{\partial A}\right)\delta A\, dx\, dy + \int \left[\int \left(\frac{\partial F}{\partial A_x}\right)\left(\frac{\partial}{\partial x}\right)(\delta A)\, dx\right] dy$$

$$+ \int \left[\int \left(\frac{\partial F}{\partial A_y}\right)\left(\frac{\partial}{\partial y}\right)(\delta A)\, dy\right] dx \qquad (21.4)$$

In Eq. (21.4), for the bracketed term in the second integral, the integration is with respect to x, keeping y constant, which implies that $dy = 0$ in Eq. (21.3), and so for this condition, we get

$$d(\delta A) = \left(\frac{\partial}{\partial x}\right)(\delta A)\, dx$$

and for the third integral, the bracketed term is integration with respect to y keeping x constant, which reduces Eq. (21.3) to

$$d(\delta A) = \left(\frac{\partial}{\partial y}\right)(\delta A)\, dy$$

So Eq. (21.4) becomes

$$\delta\Gamma = \iint \left(\frac{\partial F}{\partial A}\right)\delta A\, dx\, dy + \int \left[\int \left(\frac{\partial F}{\partial A_x}\right) d(\delta A)\right] dy + \int \left[\int \left(\frac{\partial F}{\partial A_y}\right) d(\delta A)\right] dx \qquad (21.5)$$

The second and the third integrals can be integrated by parts, i.e.

$$\int \left[\int \left(\frac{\partial F}{\partial A_x} \right) d(\delta A) \right] dy = \int \left[\left(\frac{\partial F}{\partial A_x} \right) \delta A - \int \delta A \, d\left(\frac{\partial F}{\partial A_x} \right) \right] dy \tag{21.6}$$

But

$$d\left(\frac{\partial F}{\partial A_x} \right) = \left(\frac{\partial}{\partial x} \right)\left(\frac{\partial F}{\partial A_x} \right) dx + \left(\frac{\partial}{\partial y} \right)\left(\frac{\partial F}{\partial A_x} \right) dy$$

$$= \left(\frac{\partial}{\partial x} \right)\left(\frac{\partial F}{\partial A_x} \right) dx$$

since, as before $dy = 0$ as this integration is with respect to x.

Substituting in Eq. (21.6),

$$\int \left[\int \left(\frac{\partial F}{\partial A_x} \right) d(\delta A) \right] dy = \int \left(\frac{\partial F}{\partial A_x} \right) \delta A \, dy - \iint \delta A \left(\frac{\partial}{\partial x} \right)\left(\frac{\partial F}{\partial A_x} \right) dx \, dy \tag{21.7}$$

Similarly the third term of Eq. (21.5) would reduce to

$$\int \left[\int \left(\frac{\partial F}{\partial A_y} \right) d(\delta A) \right] dx = \int \left(\frac{\partial F}{\partial A_y} \right) \delta A \, dx - \iint \delta A \left(\frac{\partial}{\partial y} \right)\left(\frac{\partial F}{\partial A_y} \right) dx \, dy \tag{21.8}$$

Substituting from Eqs. (21.7) and (21.8) in Eq. (21.5), the incremental change in the functional is obtained as

$$\delta \Gamma = \iint \left[\delta A \left(\frac{\partial F}{\partial A} \right) - \delta A \left(\frac{\partial}{\partial x} \right)\left(\frac{\partial F}{\partial A_x} \right) - \delta A \left(\frac{\partial}{\partial y} \right)\left(\frac{\partial F}{\partial A_y} \right) \right] dx \, dy$$

$$+ \left[\int \delta A \left(\frac{\partial F}{\partial A_x} \right) dy + \int \delta A \left(\frac{\partial F}{\partial A_y} \right) dx \right] \tag{21.9}$$

The above expression is called the first variation of the functional for sufficiently small values of δA. The necessary condition for the functional to attain its minimum is obtained when this variation is set to zero, i.e.

$$\delta \Gamma = 0 \tag{21.10}$$

If, in this expression, as shown in Eq. (21.9), the second (line) integral is set to zero, we are left with the first integral which also must be made equal to zero to satisfy the above condition (21.10). The implication of the line integral setting to zero will be discussed later when we discuss the boundary conditions, with the actual functional for specific problems.

For the first integral, since $\delta A \neq 0$, we are left with the following equation:

$$\delta \Gamma = \left(\frac{\partial F}{\partial A} \right) - \left(\frac{\partial}{\partial x} \right)\left(\frac{\partial F}{\partial A_x} \right) - \left(\frac{\partial}{\partial y} \right)\left(\frac{\partial F}{\partial A_y} \right) = 0 \tag{21.11}$$

This is called the Euler equation of the functional Γ. Thus the minimization of the functional requires that the Euler equation of the functional be satisfied along with the boundary integral being set to zero. Thus, if for a given functional, the above equation turns out to be the original partial differential equation which we seek to solve and force the boundary conditions so as to

make the line integrals to be zero, then the minimization of the functional will yield the solution of the partial differential equation. So now we shall select the functional for electromagnetic eddy current problems with time-harmonic variations.

The appropriate functional for solving Eq. (21.11) for the type of electromagnetic field, as specified just above, is one containing all energy terms expressed as functions of the variables which are the magnetic vetor potential and the space (i.e. the coordinate variables). The required functional for two-dimensional electromagnetic eddy-current, time-harmonic field problems is

$$\Gamma = \int \left[\int_R \left(\frac{1}{2}\right)(\nu B\, dB) - (\mathbf{J} \cdot \mathbf{A})\, dA + \left(\frac{1}{2}\right)(j\omega\sigma A^2) \right] dS \tag{21.12}$$

where

R is the region of analysis
ν is the reluctivity (reluctance)
$\mathbf{J}$ is the current density
B is the flux density
$\mathbf{A}$ is the magnetic vector potential
ω is the angular frequency
σ is the conductivity of the medium.

The first term represents the stored energy in the system; whilst the second term denotes the input energy to the system. The third term represents the reaction of the induced field; i.e. the losses in the system. Thus the functional represents the energy balance of the complete system. If the variable vector potential is changed by a small value, there will be a corresponding resulting change in the functional. This change will be 'zero' only if the functional is at its minimum. If the medium is linear, then the functional can be simplified to:

$$\Gamma = \iint \left[\left(\frac{1}{2}\right)(\nu B^2) - (\mathbf{J} \cdot \mathbf{A}) + \left(\frac{1}{2}\right)(j\omega\sigma A^2) \right] dS$$

$$= \iint \left[\left(\frac{1}{2}\right)\nu(B_x^2 + B_y^2) - (\mathbf{J} \cdot \mathbf{A}) + \left(\frac{1}{2}\right)(j\omega\sigma A^2) \right] dS$$

Writing $\mathbf{B}$ in terms of the vector potential $\mathbf{A}$, i.e.

$$B_x = \frac{\partial A}{\partial y} = A_y, \quad B_y = -\frac{\partial A}{\partial x} = -A_x$$

$$\therefore \quad \Gamma = \iint \left[\left(\frac{1}{2}\right)\nu(A_x^2 + A_y^2) - (\mathbf{J} \cdot \mathbf{A}) + \left(\frac{1}{2}\right)(j\omega\sigma A^2) \right] dS \tag{21.13}$$

The Euler equation of the above functional is given by

$$\left(\frac{\partial F}{\partial A}\right) - \left(\frac{\partial}{\partial x}\right)\left(\frac{\partial F}{\partial A_x}\right) - \left(\frac{\partial}{\partial y}\right)\left(\frac{\partial F}{\partial A_y}\right) = 0 \tag{21.14}$$

where

$$F = \left[\left(\frac{1}{2}\right)\nu(A_x^2 + A_y^2) - (\mathbf{J} \cdot \mathbf{A}) + \left(\frac{1}{2}\right)(j\omega\sigma A^2) \right] \tag{21.15}$$

Substituting for F in the Euler equation, and differentiating

$$(-J + j\omega\sigma A) - \nu\left(\frac{\partial}{\partial x}\right)A_x - \nu\left(\frac{\partial}{\partial y}\right)A_y = 0$$

or

$$(-J + j\omega\sigma A) - \nu\left(\frac{\partial}{\partial x}\right)\left(\frac{\partial A}{\partial x}\right) - \nu\left(\frac{\partial}{\partial y}\right)\left(\frac{\partial A}{\partial y}\right) = 0$$

or

$$\frac{\partial^2 A}{\partial x^2} + \frac{\partial^2 A}{\partial y^2} = -\mu(J + j\omega\sigma A) \tag{21.16}$$

This is the governing equation for two-dimensional field problems expressed in terms of the magnetic vector potential. Thus the minimization of the functional indirectly gives the solution of the equation itself.

The line-integral of the equation is given by

$$\int \delta A\left(\frac{\partial F}{\partial A_x}\right)dy + \int \delta A\left(\frac{\partial F}{\partial A_y}\right)dx = 0$$

Substituting for F in Eq. (21.15),

$$\int \delta A\, A_x\, dy + \int \delta A\, A_y\, dx = 0$$

or

$$\int \delta A\left(\frac{\partial A}{\partial x}\right)dy + \int \delta A\left(\frac{\partial A}{\partial y}\right)dx = 0$$

Replacing dx and dy by ($ds \cos \alpha$) and ($ds \sin \alpha$) respectively,

$$\int \left[\left(\frac{\partial A}{\partial x}\right)\sin\alpha + \left(\frac{\partial A}{\partial y}\right)\cos\alpha\right]\delta A\, ds = 0$$

Now, we have

$$\left(\frac{\partial A}{\partial n}\right) = \left(\frac{\partial A}{\partial x}\right)\left(\frac{dx}{dn}\right) + \left(\frac{\partial A}{\partial y}\right)\left(\frac{dy}{dn}\right)$$

However,

$$\frac{dx}{dn} = \sin\alpha \quad \text{and} \quad \frac{dy}{dn} = \cos\alpha$$

$\therefore$ We get

$$\int \left[\left(\frac{\partial A}{\partial n}\right)\delta A\right] ds = 0 \tag{21.17}$$

The above condition derived by Eq. (21.17) is satisfied on the boundary, if either of the following statements given below is true:

1. $\delta A = 0$. This makes the boundary an equipotential line or a flux line.
2. $\dfrac{\partial A}{\partial n} = 0$. This requires the normal derivative of the vector potential to be zero, or the flux lines to cross this boundary normally.

Thus, in the process of minimization of the functional, we have obtained the Euler equation of the functional (which, of course, comes out to be the original pde), and the boundary conditions, such as the continuity of the flux density vector across the boundary, are automatically satisfied, while the specified potentials on the boundaries have to be imposed in the solution process.

To minimize the functional over the region, the variable (i.e. the vector potential over the whole region) is represented as a discrete quantity. To do this the whole region is sub-divided into a number of elements in any desired manner ensuring that all the material interfaces (i.e. all the iron–air–conducvtor boundaries) coincide with the sides of the elements. The number, shape, and size of the elements can be chosen in any arbitrary manner. An approximate distribution of the vector potential is assumed within each element, such that it is a function of the coordinates of the nodes of the element and the nodal values of the potentials at its vertices. Thus an approximate distribution of the vector potential is obtained by taking the above values together. These potentials are now varied until the functional given by Eq. (21.13) reaches its minimum. The resulting vector potential distribution gives the best possible result in the 'least square' sense, since the functional contains terms involving the second order of the vector potential. A variety of approximate descriptions for A within the element are possible. One of the simplest is triangular element with linear variation of A inside, which gives sufficiently accurate results. The treatment of such a triangular element is similar to that described in Section 5.6 with reference to electrostatic problems. The only difference is that instead of the scalar potential V of the static problems, we now use A—the vector potential, which we write as (referring to Figure 5.13):

$$A\ (x,\ y) = ax + by + c \tag{21.18}$$

The constants a, b, c can be obtained in terms of the nodal values of A and their coordinates, to give:

$$A(x,\ y) = \sum_{i=1,2,3} \left(\frac{p_i + q_i x + r_i y}{2\Delta} \right) A_i \tag{21.19}$$

where

Δ = area of the triangle, i, j, k being the vertices of the triangle

$p_i = (x_j y_k - x_k y_j)$

$q_i = (y_j - y_k)$

$r_i = (x_k - x_j)$

A_i = vector potential at the node i.

The flux density within the triangle is given by

$$B_x = \left(\frac{\partial A}{\partial y} \right) = \sum_{i=1,2,3} \left(\frac{r_i A_i}{2\Delta} \right) \tag{21.20a}$$

$$B_y = -\left(\frac{\partial A}{\partial x} \right) = -\sum_{i=1,2,3} \left(\frac{q_i A_i}{2\Delta} \right) \tag{21.20b}$$

[compare with Eq. (5.83)]

In this case, the flux density, and hence the permeability inside a triangle is constant, for the first order elements.

As in Eqs. (5.84) and (5.85), we can now express A in terms of area coordinates (or shape functions), i.e.

$$A(x, y) = \sum_{i=1,2,3} c_i A_i \tag{21.21}$$

c_i, c_j, c_k being the area coordinates for the nodes i, j, and k respectively.

$$c_i = \frac{p_i + q_i x + r_i y}{2\Delta} \tag{21.22}$$

The properties of these functions have already been described in Section 5.6, and so there is no need to repeat them here. The only difference is that the functional is now

$$\Gamma = \sum_{k=1,2,\ldots,n} \int \left[\left(\frac{1}{2} \right) \nu B_k^2 - \mathbf{J}_k \cdot \mathbf{A}_k + \left(\frac{1}{2} \right) j\omega\sigma A_k^2 \right] dS_k \tag{21.23}$$

where

n is the number of elements

dS_k is the area of the kth element.

Rest of the treatment is similar to what has been discussed earlier in Section 5.6, and so the final global matrix is now given by

$$[S]\,[A] + [T]\,[A] = -\,[J] \tag{21.24}$$

The points regarding the Dirichlet boundary condition are also same as before, and the final matrix equation, containing the set of simultaneous equations, is solved by the Gaussian Elimination method.

21.3 FINITE ELEMENT METHOD FOR EDDY CURRENT PROBLEMS

We have seen earlier (Sections 15.1, 15.4, and 15.5) that the current distribution within the section of a conductor is affected by its own field, when the current is alternating. The result is that the final distribution of the current density over the conductor cross-section is non-uniform, producing higher losses which shows up as increased apparent resistance of the conductor. Thus the effects of the induced currents cannot be neglected for evaluating the losses and the current densities correctly.

The governing equation for the fields in a linear conducting region, with the excitation having sinusoidal time-variation is

$$\nabla^2 A - j\omega\mu\sigma A = -\,\mu J_s$$

or

$$\nabla^2 A = -\,\mu\,(J_s - j\omega\sigma A) \tag{21.25}$$

The first term on the right-hand side is the uniform source current density within the section of the conductor, whilst the second term represents the effects of the induced currents, or the reaction field due to the eddy currents. More generally, the second term reflects the combined effects of the boundaries and other conductors (i.e. combined skin-effects as well as proximity effects). Thus the solution obtained corresponds to the current density distribution given by the right-hand side of Eq. (21.25), or for the total current which is

$$I' = \int (J_s - j\omega\sigma A)\, dS \tag{21.26}$$

For a linear case, subject to the same boundary conditions, the resultant distribution of A, obtained, remains unchanged, except for the magnitude and the phase for any other value of the total current in that conductor. Hence the field distribution for any specified current I in a single conductor, the final value of A is given by

$$A = A'\left(\frac{I}{I'}\right) \tag{21.27}$$

where

A' is the vector potential values obtained by feeding an arbitrary current I' (corresponding to a uniform direct current density J_s)

I' is the resultant final current given by Eq. (21.26)

I is the specified current in the conductor.

The current density at any point within the conductor section is

$$J = (J_s - j\omega\sigma A) \tag{21.28}$$

where J_s is the uniform source current density.

The total current is

$$I = \int (J_s - j\omega\sigma A)\, dS \tag{21.29}$$

It should be noted that all the quantities like A, J_s, etc. are complex.

21.3.1 Single Circular Conductor in Open Space

We apply the above method to obtain the solution for a single circular conductor carrying an alternating current, and hence having induced eddy currents in it. An analytical, closed-form solution exists for this problem (Sections 15.5 and 15.5.1), and hence has been chosen to show the effectiveness of the F.E. method, and compare the accuracy of the F.E. results with the numerical values obtained from the analytical solution. The results are presented graphically in Figure 21.1,

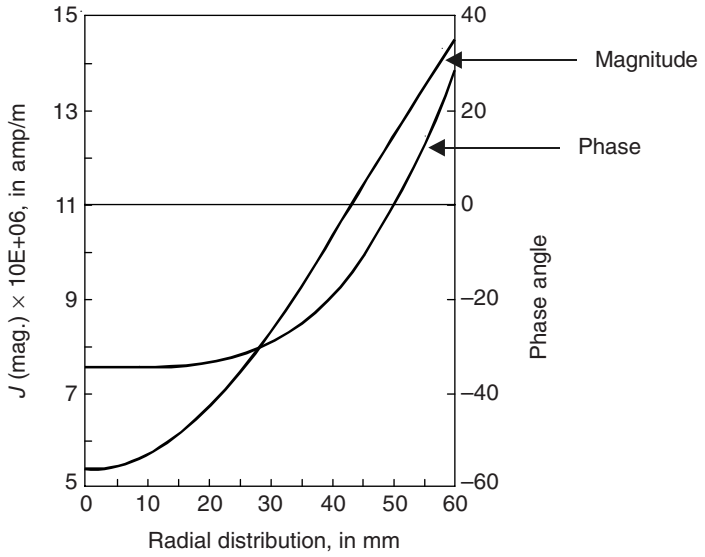


Figure 21.1 Current density distribution inside a circular conductor.

and numerically tabulated in Table 21.1. Figure 21.1 shows the current density amplitudes and its phases presented as a function of the radial distance from the centre of the circular conductor. Graphically, the results obtained by the two methods are practically coincident; and as shown in Table 21.1, the numercal values obtained, agree remarkably for the first two places of decimals, which can be considered to be very good.

Table 21.1 Current density distribution inside a circular conductor

Radius of conductor = 60.0 mm, current = 10000.0 amps
Conductivity = 5.8E+07, Frequency = 50.0 Hz
Average current density = 0.885118E+07 amps/m^2

		Analytical solution		*FEM solution*	
Sr. No.	*Radius*	*J (mag.)*	*Phase*	*J (mag.)*	*Phase*
1	0.0	0.756377E+07	–56.08	0.757594E+07	–56.00
2	8.0	0.756630E+07	–53.98	0.757869E+07	–53.90
3	12.0	0.756661E+07	–51.36	0.758918E+07	–52.80
4	16.0	0.760428E+07	–47.70	0.761703E+07	–47.70
5	20.0	0.766237E+07	–43.04	0.767522E+07	–43.00
6	24.0	0.776708E+07	–37.41	0.777990E+07	–37.90
7	28.0	0.793702E+07	–30.91	0.794965E+07	–30.90
8	32.0	0.819206E+07	–23.66	0.820429E+07	–23.70
9	36.0	0.855192E+07	–15.83	0.856353E+07	–15.80
10	40.0	0.903479E+07	–7.58	0.904555E+07	–7.60
11	42.0	0.932738E+07	–3.36	0.933765E+07	–3.40
12	44.0	0.965642E+07	0.91	0.966614E+07	0.90
13	46.0	0.100235E+08	5.19	0.100326E+08	5.20
14	48.0	0.104299E+08	9.45	0.104385E+08	9.50
15	50.0	0.108772E+08	13.78	0.108851E+08	13.70
16	52.0	0.113665E+08	18.05	0.113737E+08	18.00
17	54.0	0.118991E+08	22.30	0.119057E+08	22.30
18	56.0	0.124763E+08	26.52	0.124822E+08	26.50
19	58.0	0.130994E+08	30.72	0.131046E+08	30.70
20	60.0	0.137699E+08	34.88	0.137744E+08	34.90

DC resistance of the bar = 1.5244E–05 Ω AC resistance of the bar = 1.9486E–05 Ω

21.3.2 Parallel, Rectangular Busbars (with Eddy Currents) Inside a Slot

Two rectangular conductors are placed inside an open (at the top) rectangular slot whose side-walls (as also the bottom floor) are infinitely permeable. The open top boundary of the slot is assumed to be a flux line (Figure 21.2). When only the bottom conductor in the slot (i.e. the conductor 1 in Figure 21.2) is carrying a current, the boundary constraints force all the flux lines to be parallel to the top boundary, and the top surface of the conductor 1. There is no field below the conductor 1 due to its own current (cf. check by the method of images; and also up to this stage, this is a problem of skin effect with certain boundary conditions imposed in the finite region). When a current is passed through the top conductor in the slot (i.e. the conductor 2 in Figure 21.2), the above described field gets modified in a manner which depends on the magnitude and the phase of the current in the conductor 2 (the current distributions in the two

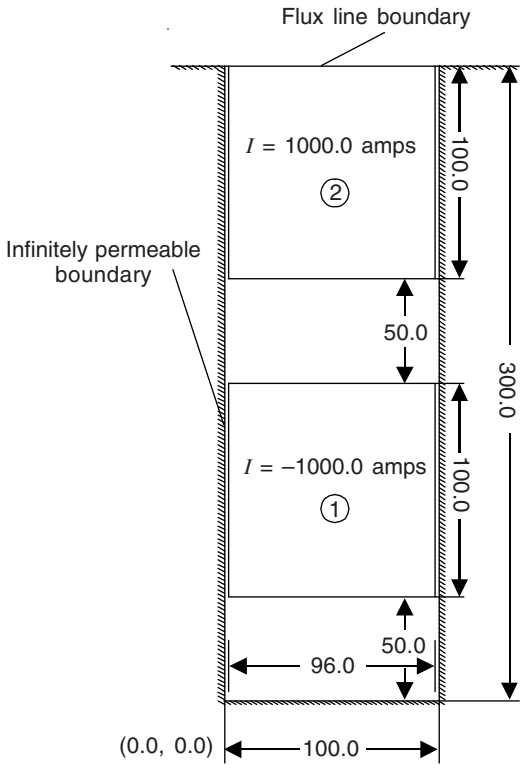


Figure 21.2 Conductors in a slot.

conductors are now a consequence of the skin effect as well as the proximity effect; and of course the superimposed effects of the boundaries in the finite region). The current density **J** and the flux density **B** distributions can be obtained by analytical method as well as the F.E.M. described earlier in this chapter. For the analytical solution, the interested reader is recommended to refer to *Foundations of Electrical Engineering* by K. Simonyi, for the solution of a similar problem of a rectangular conductor in an open rectangular slot. The current density and the flux density distributions for the top conductor in the slot have been ontained by both the analytical method as well as by the F.E.M. The results of the two methods have been compared in Figure 21.3 (i.e. the current density magnitudes at different vertical heights in the slot and the corresponding phase of the current in the top conductor in the slot), and tabulated in Table 21.2, for equal and opposite currents in the two conductors. The agreement between the two sets of results is excellent, indicating the validity of the method. Also, the magnetic flux is confined to the space between the top of the top conductor (i.e. the conductor 2) and the bottom of the lower conductor (i.e. the conductor 1). The resultant field is shown in Figure 21.4. For the calculations, the dimensions are as indicated in Figure 21.2, and the conductivity σ of the conductors = 5.8×10^6 mhos/metre.

For the top conductor, it is found that

$$\frac{R_{ac}}{R_{dc}} = 3.322$$

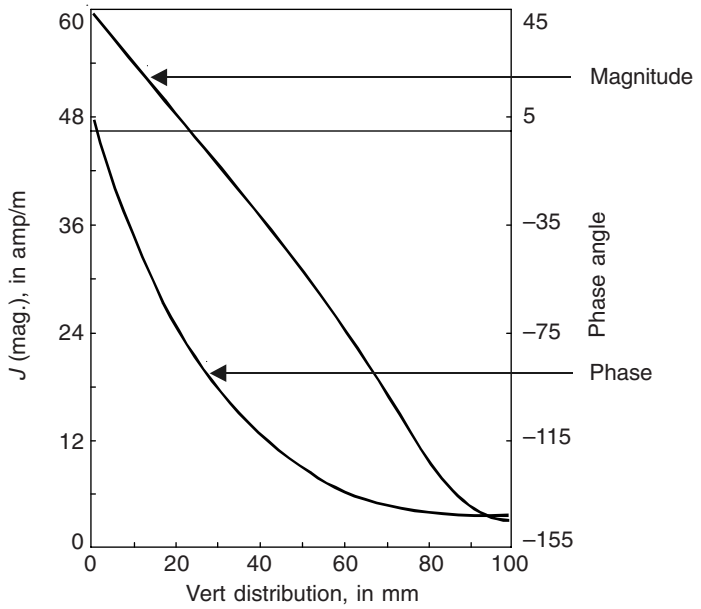


Figure 21.3 Current density distribution inside a conductor in a slot.

Table 21.2 Current density distribution inside a conductor in a slot

Conductor current = 10000.0 amps
Conductivity = 5.8E+07, Frequency = 50.0 Hz
Average current density = 0.10416E+07 amps/m^2

		Analytical solution		*FEM solution*	
Sr. No.	*Dist.*	*J (mag.)*	*Phase*	*J (mag.)*	*Phase*
1	0.8	0.47682E+07	43.43	0.47353E+07	42.90
2	4.2	0.42617E+07	36.99	0.42386E+07	36.50
3	5.8	0.40423E+07	33.96	0.40104E+07	33.20
4	9.2	0.36129E+07	27.53	0.35911E+07	26.80
5	11.7	0.33264E+07	22.82	0.33058E+07	22.20
6	18.3	0.26734E+07	10.41	0.26512E+07	9.40
7	21.7	0.23877E+07	4.04	0.23736E+07	3.00
8	28.3	0.19143E+07	−8.29	0.18997E+07	−9.60
9	34.2	0.15668E+07	−19.30	0.15594E+07	−20.90
10	39.2	0.13184E+06	−28.69	0.13129E+07	−30.40
11	44.2	0.11060E+07	−38.22	0.11021E+06	−40.00
12	49.2	0.92493E+06	−48.02	0.92238E+06	−49.80
13	54.2	0.77215E+06	−58.28	0.77079E+06	−60.20
14	59.2	0.64565E+06	−69.20	0.64543E+06	−71.10
15	68.3	0.47961E+06	−91.17	0.48033E+06	−93.10
16	78.3	0.38575E+06	−116.60	0.38820E+06	−118.20
17	84.2	0.36407E+06	−129.46	0.36681E+06	−130.80
18	89.2	0.35720E+06	−137.65	0.35942E+06	−138.60
19	94.2	0.35541E+06	−142.86	0.35673E+06	−143.40
20	99.2	0.35526E+06	−144.93	0.35540E+06	−145.00

DC resistance of the bar = 1.7959E–05 Ω AC resistance of the bar = 5.9776E–05 Ω

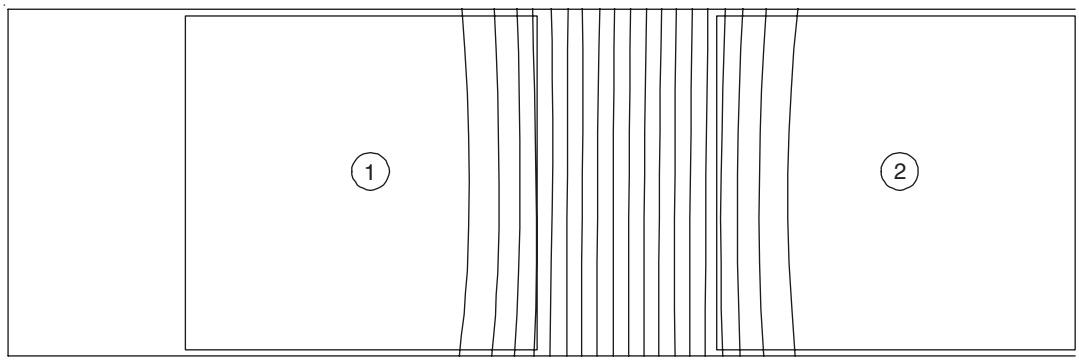


Figure 21.4 Conductor in a slot with eddy currents—FEM solution—analysis in *x*-*y* coordinates at $t = 0$.

21.3.3 Number of Parallel Busbars Enclosed in a Rectangular Conducting Box

The F.E.M. described has been used to analyze the field distribution of eight, parallel, rectangular, copper busbars enclosed in a rectangular box (whose sides are also parallel to the length dimension of the busbars), made up of a conducting material of finite resistivity and permeability. Only half the section of the box has been considered for analysis, because of the symmetry considerations. Figure 21.5 shows the half-section of the box (or the tank), containing the busbars

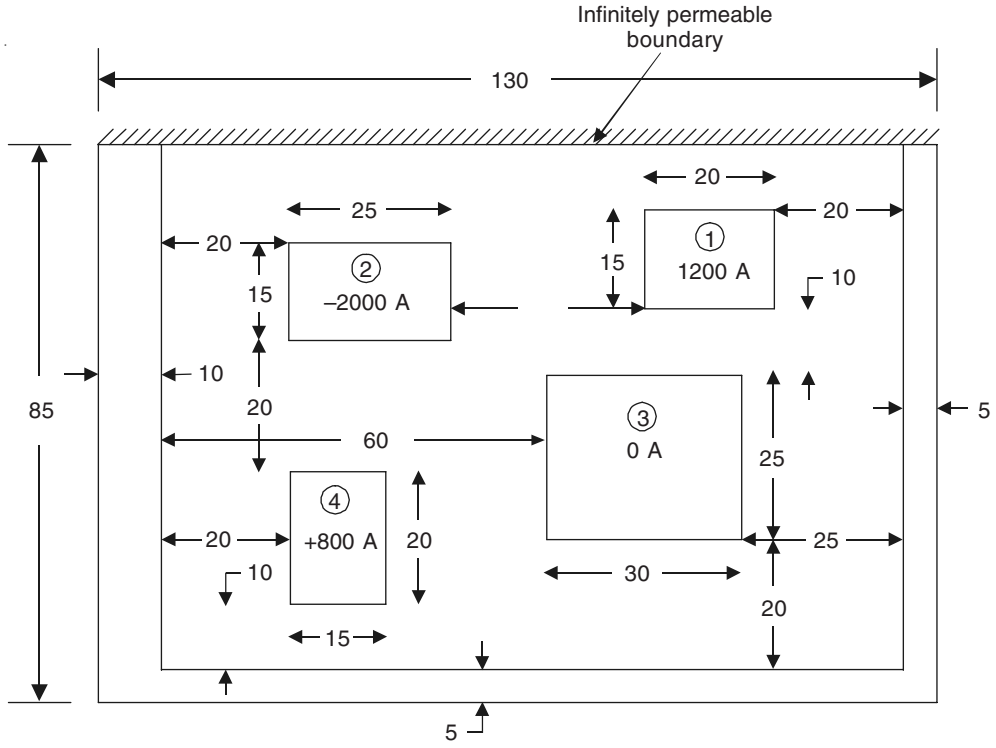


Figure 21.5 Busbars enclosed in a conducting box.

with all the dimensions and the currents indicated on the figure. It will be seen that the conductor 3 is a dummy one carrying zero current. However this busbar as well as the tank walls will have eddy currents induced in them, causing losses therein. These losses will be reflected as additional resistances in the other conductors, which would be a function of the other conductors and also of the tank walls to each conductor. For these calculations,

$$\sigma_{\text{busbars}} = 5.0 \times 10^{+7} \text{ mhos/metre}$$
$$\sigma_{\text{tank walls}} = 1.666 \times 10^{+7} \text{ mhos/metre}$$
$$\mu_{\text{r, tank wall}} = 50$$

The resultant field distribution obtained is shown in Figure 21.6 at the instant of time $t = 0$, when the currents are at their peak values. The flux distribution at the instant of time $t = \pi/2$, when the currents are passing through their zeroes is shown in Figure 21.7. This picture shows the flux due to the induced currents. When the ac resistances of the different conductors are calculated, it is found that the ratio (R_{ac}/R_{dc}) is different for the different conductors.

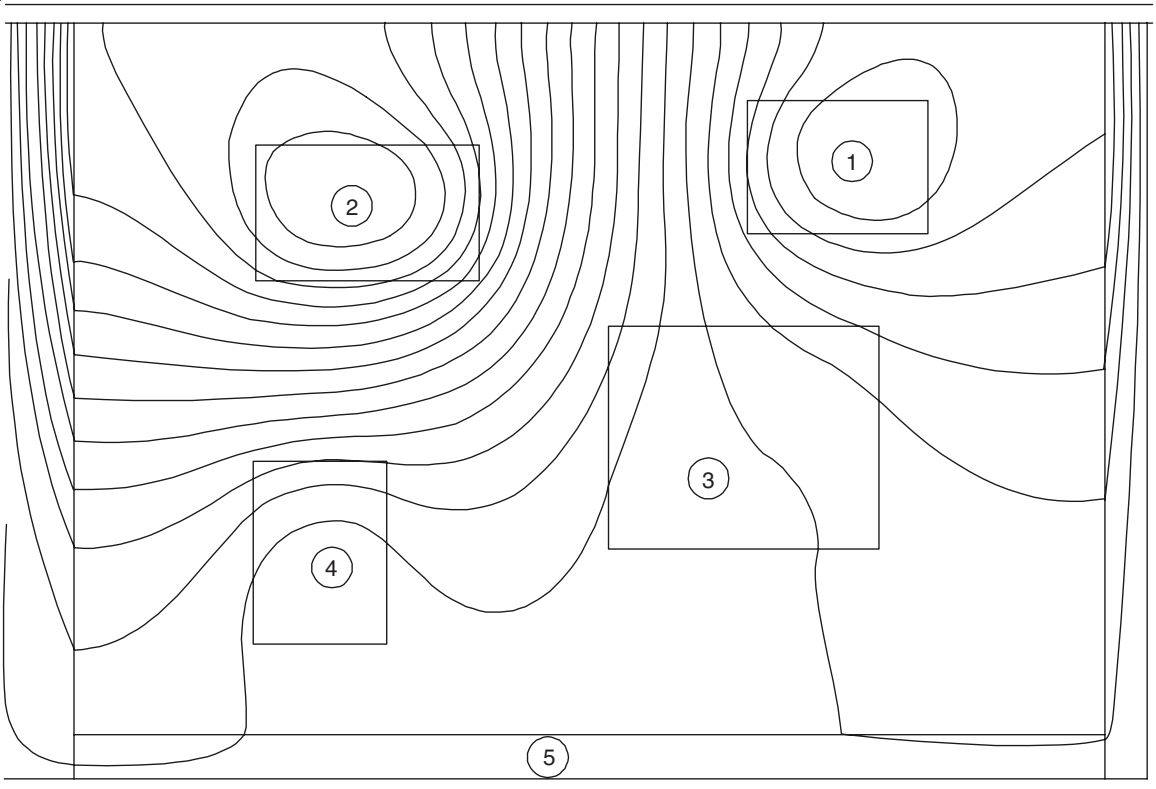


Figure 21.6 Busbars enclosed in a conducting box of mild steel—half geometry with dummy bar—analysis in *x-y* coordinates at $t = 0$.

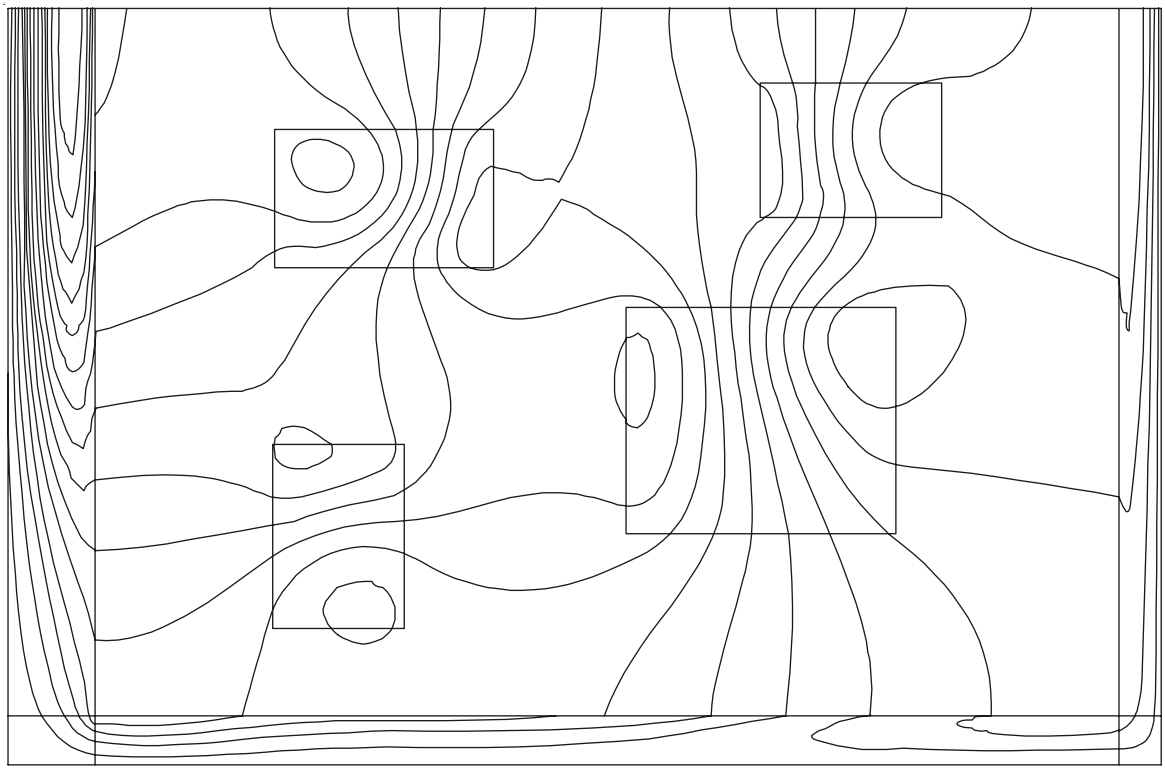


Figure 21.7 Busbars enclosed in a conducting box of mild steel—half geometry with dummy bar—analysis in x-y coordinates at $t = \frac{\pi}{2}$.

21.4 APPLICATION OF THE F.E.M. TO TRANSFORMERS AND DC MACHINES

The method of calculating the magnetic field and the eddy current distributions has been successfully applied to all types of rotating machines and other allied electrical devices. Figures 21.8 and 21.9 show the flux distributions in the two orthogonal sections of a generator transformer, and Figures 21.10 to 21.13 show the flux density distributions in the core section (of the armature, field pole and the yoke of the stator of the machine over one pole-pitch) of a DC motor, under different load conditions, i.e. 50%, 100%, 200%, and 300%. Figures 21.14 and 21.15 show the flux distributions in the rotor region of a squirrel cage induction motor. It should be noted that for the DC motor, the field has no eddy currents and its analysis has been treated as a magnetostatic field problem.

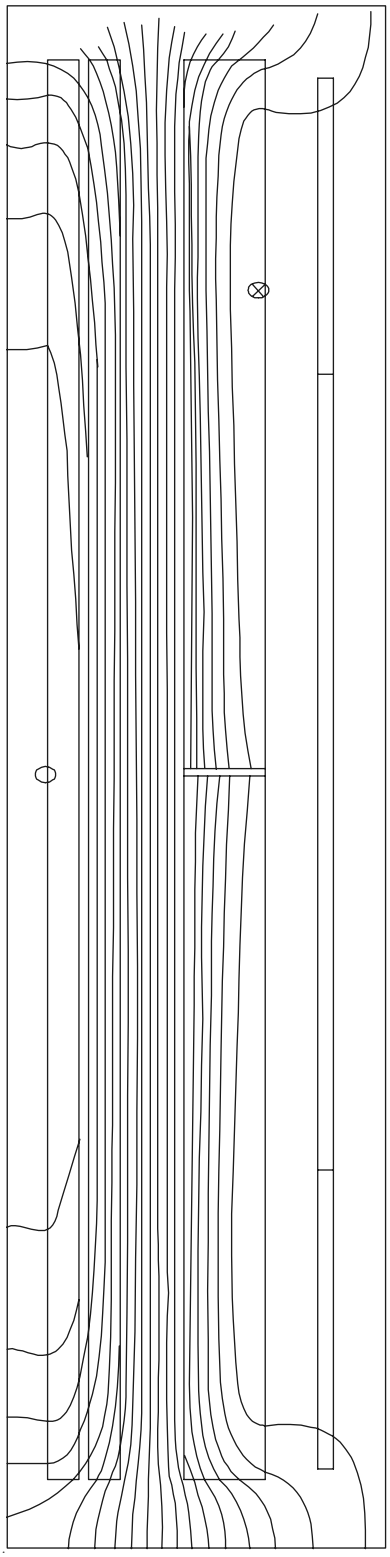


Figure 21.8 250 MVA 220 kV generator transformer window section—analysis in *r-z* coordinates.

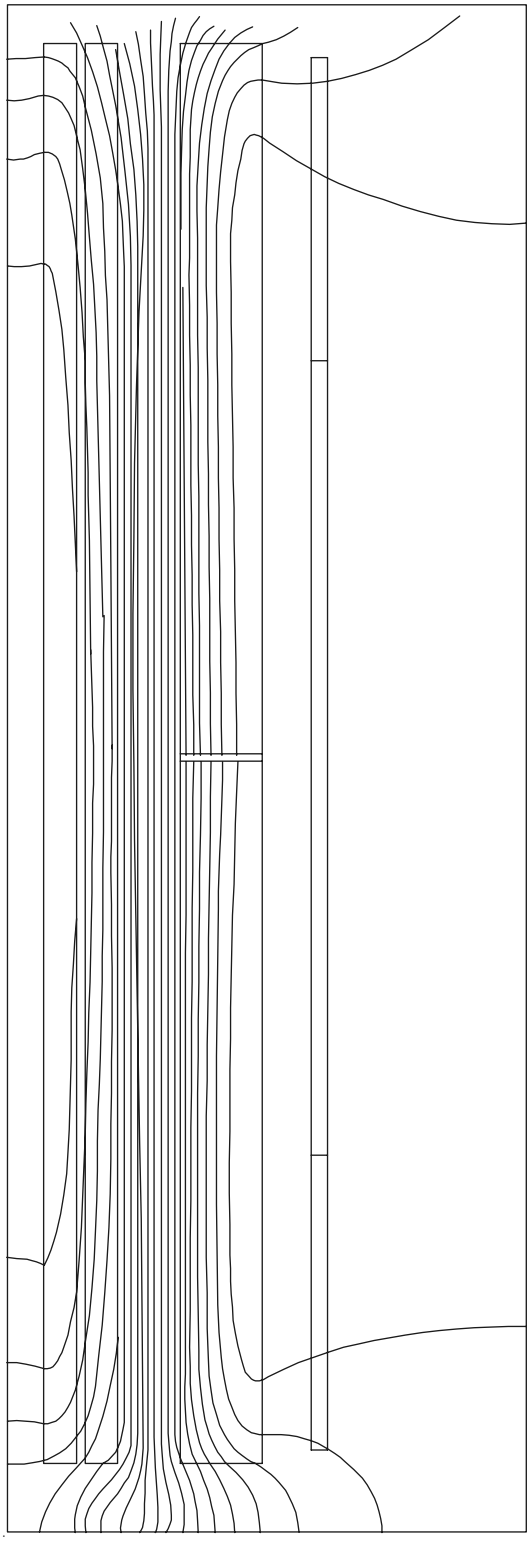


Figure 21.9 250 MVA 220 kV generator transformer—orthogonal section—analysis in *r-z* coordinates.

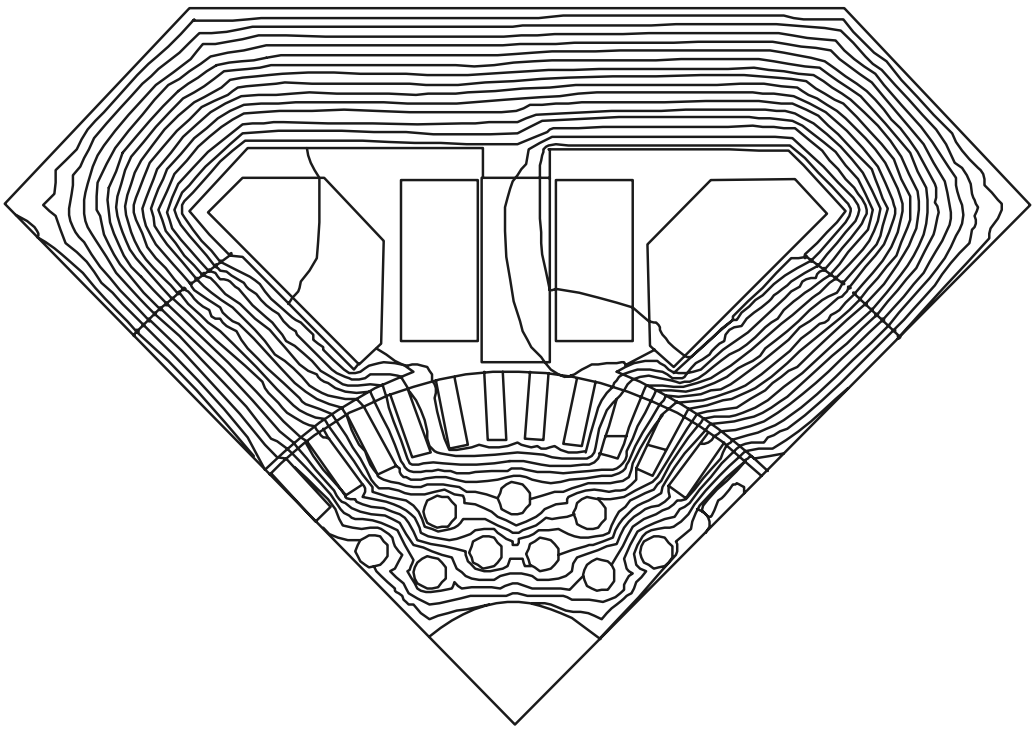


Figure 21.10 DC machine—flux distribution on 50% load.

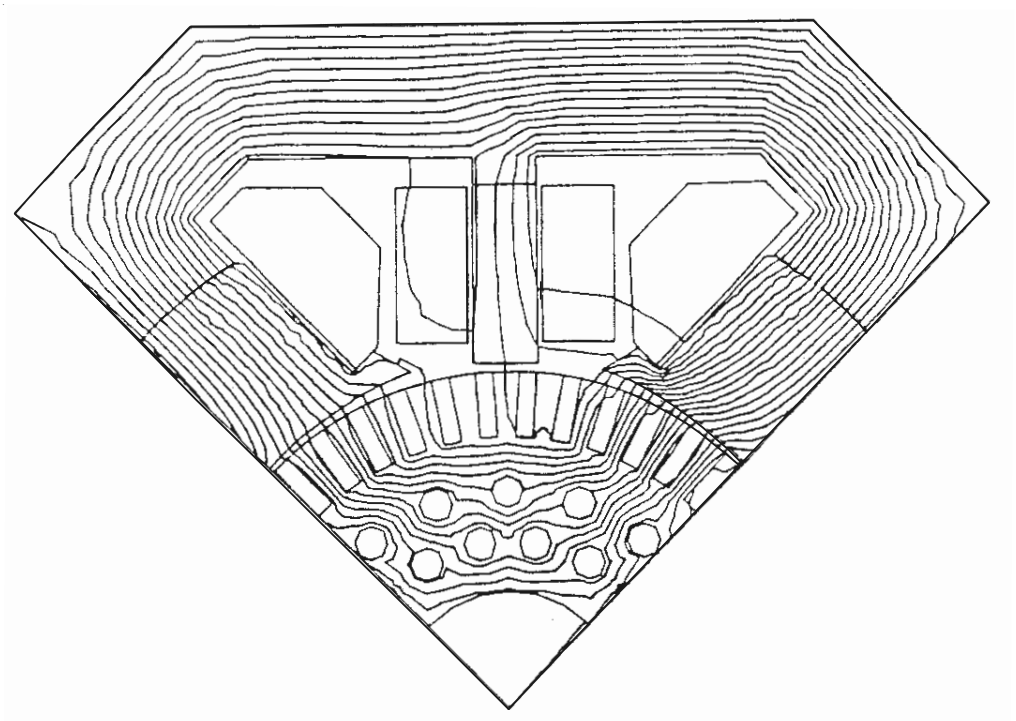


Figure 21.11 DC machine—flux distribution on 100% load.

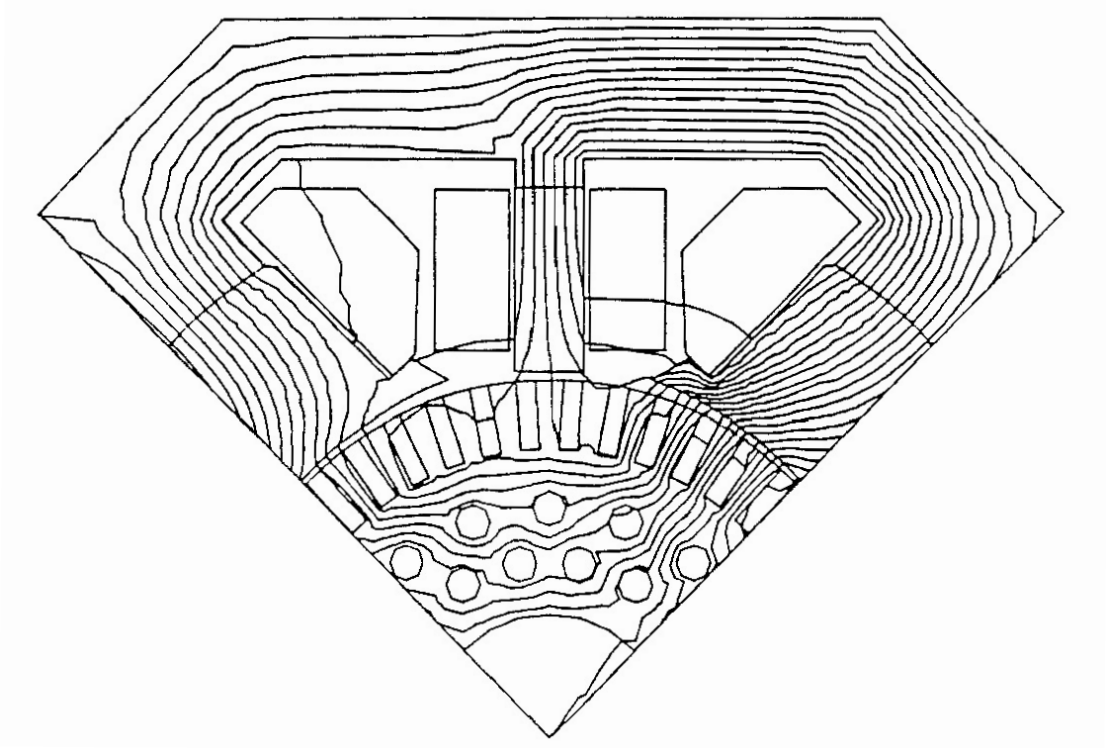


Figure 21.12 DC machine—flux distribution on 200% load.

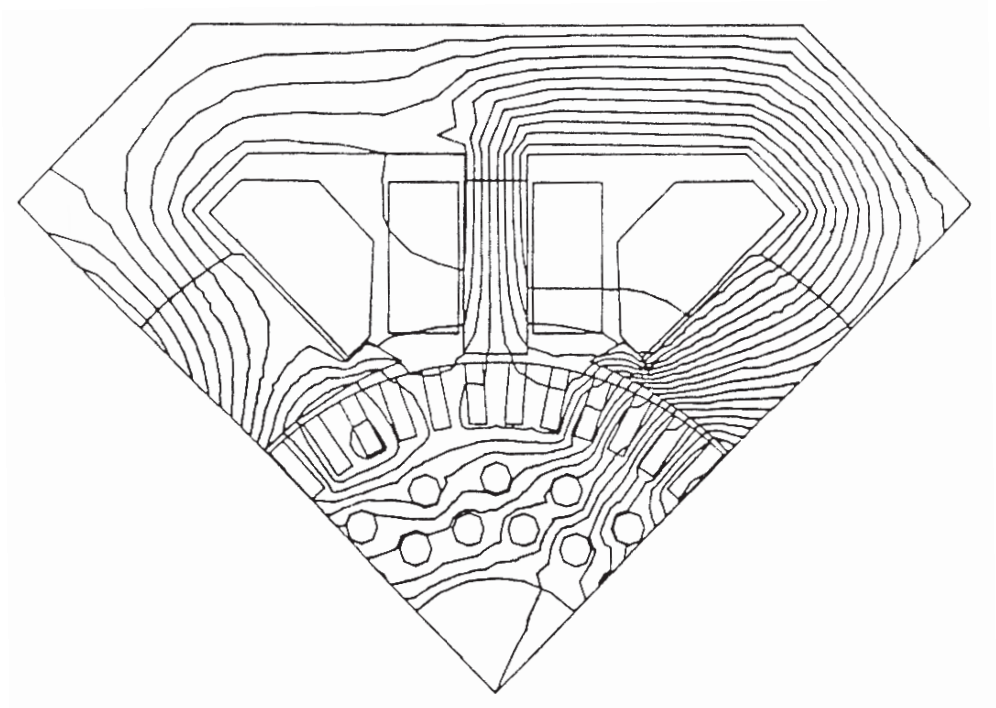


Figure 21.13 DC machine—flux distribution on 300% load.

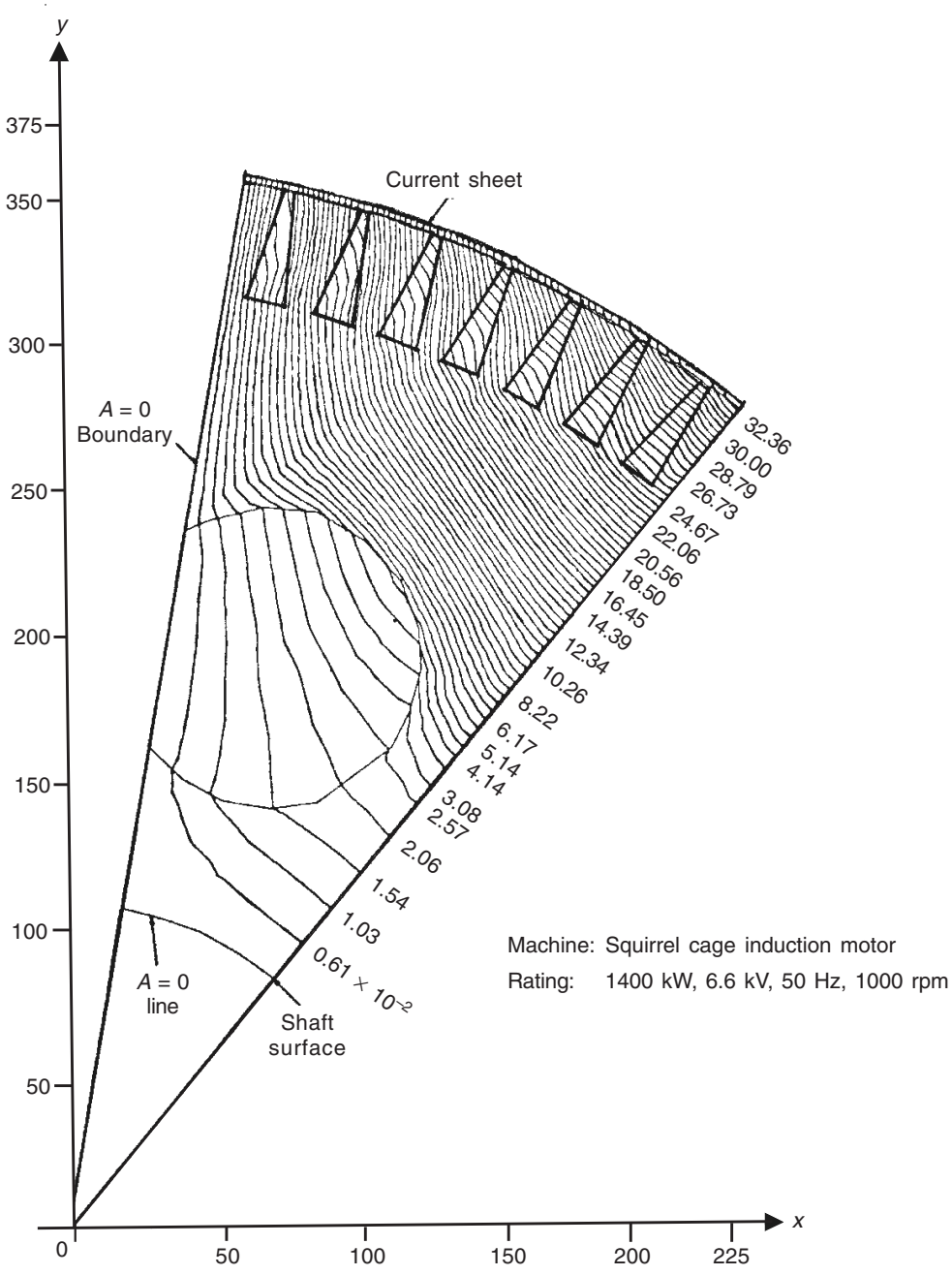


Figure 21.14 Magnetic flux distribution in the rotor region for circular ventilating hole, without shaft.

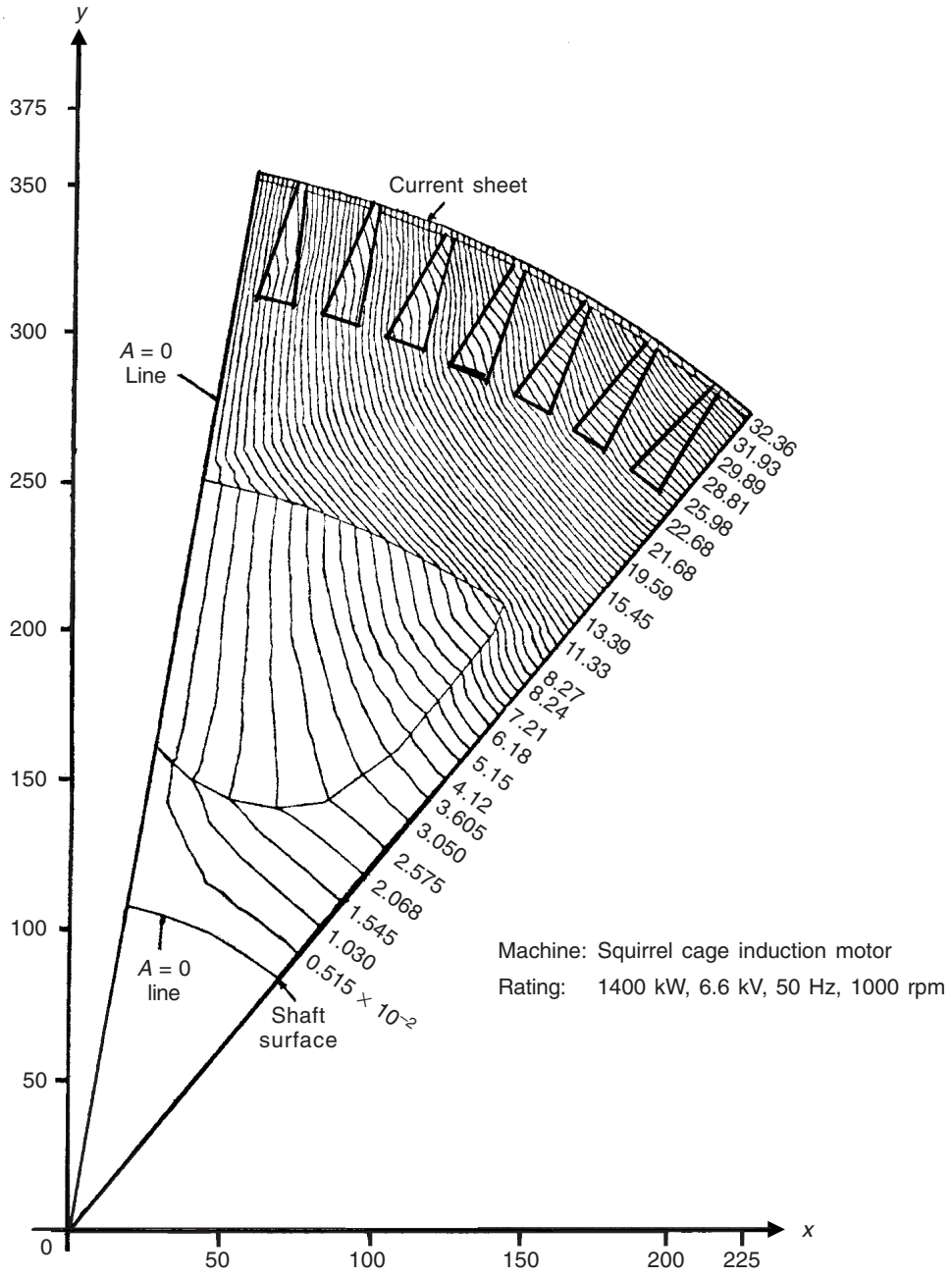


Figure 21.15 Magnetic flux distribution in the rotor region for modified shape ventilating hole, without shaft.

21.5 SOME PARTIAL DIFFERENTIAL EQUATIONS AND THEIR FUNCTIONALS

We now list some of the other very commonly met PDEs and their functionals in electromagnetic problems.

(a) Poisson's equation (scalar)

$$\nu \nabla^2 A = -J, \qquad \text{and}$$

$$F = \left(\frac{\nu}{2}\right)\int_{\Omega} |\nabla A|^2 \, d\Omega - \int_{\Omega} (\mathbf{J} \cdot \mathbf{A}) \, d\Omega$$

(b) Diffusion equation (and/or Eddy current equation)

$$\nu \nabla^2 A = j\,\omega\,\sigma A - J, \qquad \text{and}$$

$$F = \left(\frac{\nu}{2}\right)\int_{\Omega} |\nabla A|^2 \, d\Omega + \left(\frac{j\omega\sigma}{2}\right)\int_{\Omega} A^2 d\Omega - \int_{\Omega} (\mathbf{J} \cdot \mathbf{A}) \, d\Omega$$

(c) Vector Poisson's equation

$$\nabla \times \nu \nabla \times \mathbf{A} = \mathbf{J}, \qquad \text{and}$$

$$F = \left(\frac{\nu}{2}\right)\int_{\Omega} |\nabla \times A|^2 \, d\Omega - \int_{\Omega} (\mathbf{J} \cdot \mathbf{A}) \, d\Omega$$

(d) Nonlinear Poisson's equation

$$\nu(A)\, \nabla^2 A = -J, \qquad \text{and}$$

$$F = \int_{\Omega} \left[\int_{0}^{B} \nu b \, db - (\mathbf{J} \cdot \mathbf{A}) \right] d\Omega$$

21.6 GENERAL COMMENTS

We have shown only a very limited set of applications of the F.E.M. to eddy current problems. This method has also been widely used for solving electromagnetic wave problems. We have not discussed such problems here as there is quite a lot of similarity between the magnetic diffusion equation and the wave equation, and the solutions would have significant points of similarity. The main difference is due to the fact that the wave equation contains a term with double derivative of time (as distinct from the diffusion equation where the equivalent term is a single derivative of time). The consequence of this difference is that for time-harmonic excitations, the eddy current equation becomes complex (due to the imaginary j) whereas the wave equation remains one in real variables.

We have not included in the present text any solutions of three-dimensional problems. Though F.E.M. has been successfully used to solve three-dimensional static problems, its application to three-dimensional eddy current problems still remains a controversial point, and is an important topic of research.

21.7 FINITE ELEMENT METHOD (PROCEDURE DETAILS)

The finite element analysis has been so widely accepted as a successful numerical method for solving a wide range of engineering problems, that it has now become an integral part of 'Computer Aided Engineering' (CAE). It is now extensively used in the analysis and design of many complex real-life engineering systems. Though originally it started as a tool for structural analysis, its application to electromagnetic field problems (amongst many other branches as well) is very well-established. So far in our discussions (i.e. in Chapter 4 Section 5.4–5.6.5, and in this Chapter 21) we have given a strong emphasis on the 'variational' basis of this method; and we will end our present discussion of this method by describing the actual steps used in applying the F.E.M to solve a typical two-dimensional problem in electromagnetism (including electrostatics, magnetostatics, electromagnetics, or in fact any two-dimensional problem).

The main steps of the F.E. method are:

1. Choose a suitable formulation of the problem, i.e. whether to use vector or scalar potential, a combination of both, interpretation of the boundaries, etc.
2. Discretization of the region under consideration (i.e. problem domain or solution region) into a finite number of sub-regions which are called 'elements'.
3. Deriving the governing equations, for a typical 'element' in the solution region.
4. Assembling all element equations in the solution region (in the global form).
5. Solving the global linear system of equations so obtained.

21.7.1 Formulation of the Problem

This is in fact a mental process. For a two-dimensional problem, based on the physical aspects, it has to be decided whether a scalar potential or a vector potential is a better choice. If it is an electrostatic field problem, then of course it can be solved by a scalar potential. On the other hand for a magnetic field problem, if the solution region does not contain any current sources, the scalar potential can be used. But if there are current-carrying regions in the solution domain then the vector potential is an obvious choice. When the source current (volume or line) is normal to the plane surface under consideration the vector potential will have only one component (e.g. in a two-dim Cartesian geometry problem in x-y plane). When the vector potential has only one component (i.e. z–component), only then the single component of **A** is similar to the scalar potential $\{\Omega\}$. This is **not** to say that A_z is identical with Ω, it is only similar to Ω because the field vectors H_x, H_y (or B_x, B_y) are in fact $B_x = -\mu\dfrac{\partial\Omega}{\partial x}, B_y = -\mu\dfrac{\partial\Omega}{\partial y}$ from the scalar potential and $B_x = \dfrac{\partial A_z}{\partial y}$ and $B_y = -\dfrac{\partial A_z}{\partial x}$ from the vector potential. Furthermore, if the source current is a line current, then it can be accounted for in the domain only if it is normal to the plane of the domain (i.e. if it is in the z-direction for a x-y plane problem). A line current in x-y plane cannot be correctly represented in a two-dimensional x-y domain because if it is so drawn then it (the line section) represents a section of a current sheet where width is in either x-z plane or y-z plane and does not represent a line current. In fact, a line current lying in the x-y plane of the solution region would produce a field pattern which would be three-dimensional in nature and hence cannot be treated as a two-dimensional problem.

The boundaries of any two-dimensional problem would be basically straightforward, i.e. Dirichlet-type or Neumann type or Mixed type. For multi-region problems, the interface continuity conditions need some care, in case there are surface currents on the interfaces.

21.7.2 Finite Element Discretization

The solution region is sub-divided into a number of 'finite elements' as shown in Figure 21.16.

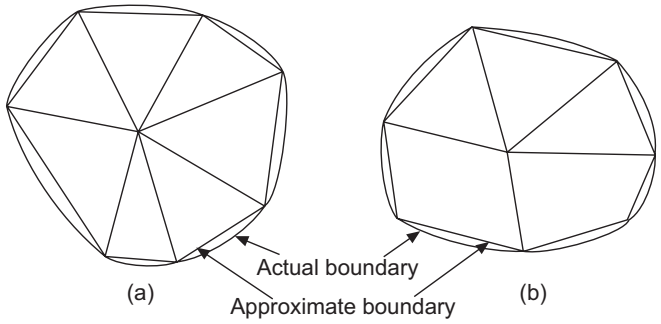


Figure 21.16 Typical irregular domain, sub-divided into (a) triangular elements only and (b) a combination of triangular and quadrilateral elements.

Region (a) has 7 triangular elements and 8 nodes.
Region (b) has 3 triangular and two quadrilateral elements and 8 nodes.
For simplicity, we consider an electrostatic field problem.

Note: If it is a multi-region problem with different material characteristics (i.e. permittivity in E.S. fields, permeability and conductivity in electromagnetic fields, etc.), then the material interface (or interfaces) should coincide with the element sides. This process ensures the continuity of normal boundary conditions across the interface.

We start by taking an approximate value for the potential in a typical element of the domain. Let the potential be V_e (say) within the element e, and then inter-relate the potential distribution in various elements, such that the potential is continuous across the inter-element boundaries. The approximate solution for the potential in the region is then,

$$V(x, y) = \sum_{e=1}^{N} V_e(x, y) \tag{21.30}$$

where

N = number of elements into which the whole domain has been subdivided

V_e = approximate potential within the element e.

It should be clearly understood that the point (x, y) in the element is 'any point' in the element e and is not restricted to a specific point. This is one of the important differences between the F.E.M and the F.D.M. (i.e. finite difference method). In the F.D.M, once the region has been discretized (say, for example, the rectangular mesh), we deal with the potential values at the 'nodes only' and the inter-nodal spaces are ignored. This is very obvious when the P.D.E under consideration is reduced to the corresponding algebraic equation by using the 'Taylor series

expansion method'. But when the 'line integral method' is used to derive the discretized equation, the implicit assumption is that the potential values vary linearly along the mesh arms.

The simplest and the most common form of approximation for V_e in a triangular element is the first-order polynomial approximation, i.e.

$$V_e(x, y) = a + bx + cy \tag{21.31}$$

And for a quadrilateral element, the corresponding approximation is

$$V_e(x, y) = a + bx + cy + dxy \tag{21.32}$$

The potential V_e is non-zero within the element e and is zero outside e. It is difficult to approximate the boundary of the problem domain by quadrilateral elements; and for the present problem we have used triangular elements only.

The assumption of linear variation of the potential within the triangular element (as given by Eq. (21.31) is equivalent to assuming uniform electric field within the element, i.e.

$$\mathbf{E}_e = -\nabla V_e = -\left(\mathbf{i}_x \frac{\partial V_e}{\partial x} + \mathbf{i}_y \frac{\partial V_e}{\partial y}\right)$$

$$= -\mathbf{i}_x b - \mathbf{i}_y c \tag{21.33}$$

On the other hand, for the quadrilateral element, it amounts to saying that the components of $\mathbf{E}_e$ vary linearly within the element, i.e.

$$E_e = -\nabla V_e$$

$$= \mathbf{i}_x (b + dy) - \mathbf{i}_y (c + dx) \tag{21.34}$$

21.7.3 Element Governing Equation

For a typical triangular element, as shown in Figure 21.17, the potentials V_{e1}, V_{e2}, and V_{e3} at the nodes 1, 2, and 3 respectively are given by [using the first-order polynomial relation of Eq. (21.31)]

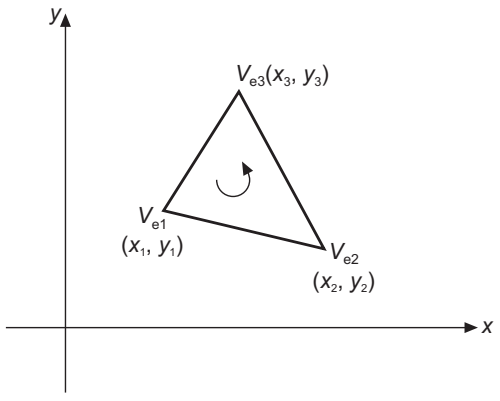


Figure 21.17 A typical triangular element.

(The local node numbering sequence 1–2–3 must be in the counterclockwise sense.)

$$\begin{bmatrix} V_{e1} \\ V_{e2} \\ V_{e3} \end{bmatrix} = \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} \tag{21.35}$$

written in compact matrix form.

The coefficients a, b, c of the polynomial representation can be determined by inverting the 3×3 matrix above, i.e.

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{bmatrix}^{-1} \begin{bmatrix} V_{e1} \\ V_{e2} \\ V_{e3} \end{bmatrix} \tag{21.36}$$

$$= \frac{1}{2\Delta} \begin{bmatrix} x_2y_3 - x_3y_2 & x_3y_1 - x_1y_3 & x_1y_2 - x_2y_1 \\ y_2 - y_3 & y_3 - y_1 & y_1 - y_2 \\ x_3 - x_2 & x_1 - x_3 & x_2 - x_1 \end{bmatrix} \begin{bmatrix} V_{e1} \\ V_{e2} \\ V_{e3} \end{bmatrix}$$

$$= \frac{1}{2\Delta} \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} \begin{bmatrix} V_{e1} \\ V_{e2} \\ V_{e3} \end{bmatrix} \tag{21.37}$$

where

$$\begin{aligned} a_1 &= x_2y_3 - x_3y_2 & a_2 &= x_3y_1 - x_1y_3 & a_3 &= x_1y_2 - x_2y_1 \\ b_1 &= y_2 - y_3 & b_2 &= y_3 - y_1 & b_3 &= y_1 - y_2 \\ c_1 &= x_3 - x_2 & c_2 &= x_1 - x_3 & c_3 &= x_2 - x_1 \end{aligned}$$

and Δ is the area of the triangular element e, i.e.

$$\Delta = \frac{1}{2} \begin{vmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{vmatrix}$$

$$= \frac{1}{2}[(x_2y_3 - x_3y_2) + (x_3y_1 - x_1y_3) + (x_1y_2 - x_2y_1)]$$

$$= \frac{1}{2}\left[(x_2 - x_1)(y_3 - y_1) - (x_3 - x_1)(y_2 - y_1)\right] = \frac{1}{2}(b_2c_3 - b_3c_2)$$

$$= \frac{1}{2}[(y_1 - y_2)(x_3 - x_2) - (y_2 - y_1)(x_2 - x_1)] = \frac{1}{2}(b_3c_1 - b_1c_3)$$

$$= \frac{1}{2}\left[(y_2 - y_3)(x_1 - x_3) - (y_3 - y_1)(x_3 - x_2)\right] = \frac{1}{2}(b_1c_2 - b_2c_1) \tag{21.38}$$

Substituting the values of a, b, c from Eq. (21.36) or (21.37) into Eq. (21.31), and writing it in matrix form:

$$V_e = [1 \quad x \quad y] \frac{1}{2\Delta} \begin{bmatrix} x_2y_3 - x_3y_2 & x_3y_1 - x_1y_3 & x_1y_2 - x_2y_1 \\ y_2 - y_3 & y_3 - y_1 & y_1 - y_2 \\ x_3 - x_2 & x_1 - x_3 & x_2 - x_1 \end{bmatrix} \begin{bmatrix} V_{e1} \\ V_{e2} \\ V_{e3} \end{bmatrix}$$

This can be written in further compact form as

$$V_e(x, y) = \sum_{i=1}^{3} N_i(x, y)\, V_{ei} \tag{21.39}$$

where N_i are the interpolation (or shape) functions used to interpolate the values of the field variable at an interior point in the element, from its value at certain key points (i.e. nodes) in the element, and are given by:

$$N_1(x, y) = \frac{1}{2\Delta}\left[(x_2y_3 - x_3y_2) + (y_2 - y_3)x + (x_3 - x_2)y\right]$$

$$= \frac{1}{2\Delta}(a_1 + b_1x + c_1y) \tag{21.40a}$$

$$N_2(x, y) = \frac{1}{2\Delta}(a_2 + b_2x + c_2y) \tag{21.40b}$$

$$N_3(x, y) = \frac{1}{2\Delta}(a_3 + b_3x + c_3y) \tag{21.40c}$$

Expanding Eq. (21.39),

$$V_e(x, y) = \frac{1}{2\Delta}\left[(a_1 + b_1x + c_1y)V_{e1} + (a_2 + b_2x + c_2y)V_{e2} + (a_3 + b_3x + c_3y)V_{e3}\right]$$

Also, it can be checked,

$$N_i(x_j, y_j) = \delta^i_j = \begin{cases} 1, & i = j; \\ 0, & i \neq j. \end{cases} \tag{21.41}$$

$\uparrow$

Kronecker Delta

[Check:

$i = j = 1$

$$N_1(x_1, y_1) = \frac{1}{2\Delta}\left[(x_2y_3 - x_3y_2) + (y_2 - y_3)x_1 + (x_3 - x_2)y_1\right]$$

$$= \frac{2\Delta}{2\Delta} = 1$$

and $(i = 1) \neq (j = 2)$

$$N_1(x_2, y_2) = \frac{1}{2\Delta}\left[(x_2y_3 - x_3y_2) + (y_2 - y_3)x_2 + (x_3 - x_2)y_2\right]$$

$$= 0]$$

Since the shape functions N_i satisfy the Kronecker Delta function, it follows that the sum of the shape functions for the triangular element would equal unity, i.e.

$$\sum_{i=1}^{3} N_i(x, y) = 1 \tag{21.41a}$$

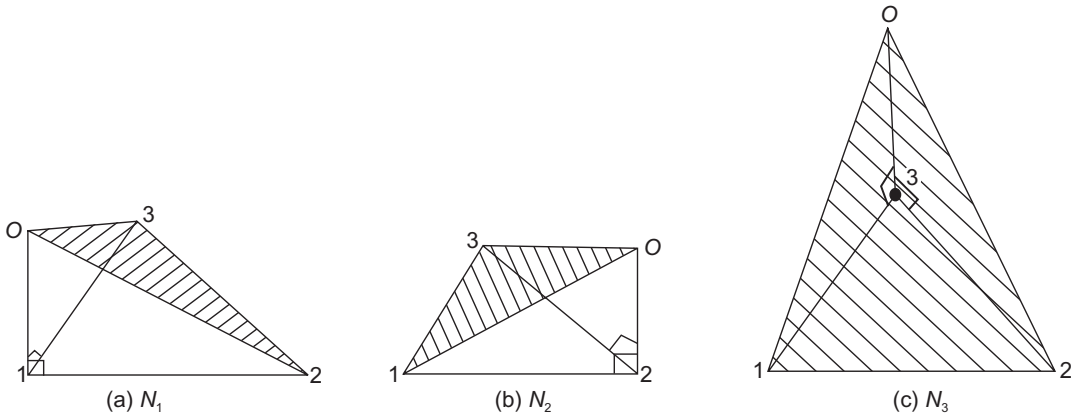


Figure 21.18 Shape functions (or linear interpolation functions) for the triangular element 1–2–3.

Note: Length $O1$, $O2$, $O3$ are unit lengths normal to the plane of the triangular element 1–2–3. Geometrically, the element triangle (1–2–3) is the projection of the shape function triangles on the x-y plane.

Next, we consider the electrostatic energy (per unit length) associated with the element under consideration (i.e. the element e) is given by Eq. (3.4) as:

$$W_e = \frac{1}{2}\iint \varepsilon |\mathbf{E}|^2 \, ds = \frac{1}{2}\iint \varepsilon |\nabla V_e|^2 dS \tag{21.42}$$

where a two-dimensional, charge-free ($\rho_0 = 0$) solution region has been assumed.

Using Eq. (21.38), we get:

$$\nabla V_e(x, y) = \sum_{i=1}^{3} V_{ei} \, \nabla N_i(x, y) \tag{21.43}$$

(Since V_{ei} are fixed potentials at the nodes of the element.)

Substituting Eq. (21.43) in Eq. (21.42),

$$W_e = \frac{1}{2}\sum_{i=1}^{3}\sum_{j=1}^{3} \varepsilon V_{ei}\left[\iint (\nabla N_i)\cdot(\nabla N_j)\, dS\right] V_{ej} \tag{21.44}$$

Defining the surface integral term in the above Eq. (21.44) as

$$C_{ij}^{(e)} = \iint (\nabla N_i)\cdot(\nabla N_j) dS \tag{21.45}$$

then W_e [in Eq. (21.44)] can be written down in the matrix form as

$$W_e = \frac{1}{2}\varepsilon [V_e]^T \, [C^{(e)}][V_e] \tag{21.46}$$

where

$$[V_e] = \begin{bmatrix} V_{e1} \\ V_{e2} \\ V_{e3} \end{bmatrix} \tag{21.47a}$$

$$[C^{(e)}] = \begin{bmatrix} C_{11}^{(e)} & C_{12}^{(e)} & C_{13}^{(e)} \\ C_{21}^{(e)} & C_{22}^{(e)} & C_{23}^{(e)} \\ C_{31}^{(e)} & C_{32}^{(e)} & C_{33}^{(e)} \end{bmatrix} \tag{21.47b}$$

↑

The element coefficient matrix

Each term (i.e. matrix element) $C_{ij}^{(e)}$ represents the coupling beween the nodes i and j. Its value can be obtained from Eq. (21.45), e.g.

$$C_{12}^{(e)} = \iint (\nabla N_1)\cdot(\nabla N_2)\,dS$$

$$= \frac{1}{2\Delta}\left[\mathbf{i}_x(y_2 - y_3) + \mathbf{i}_y(x_3 - x_2)\right]\cdot\frac{1}{2\Delta}\left[\mathbf{i}_x(y_3 - y_1) + \mathbf{i}_y(x_1 - x_3)\right]\iint dS$$

[from Eqs. (21.40) and (21.38)]

$$= \frac{1}{4\Delta^2}\left[(y_2 - y_3)(y_3 - y_1) + (x_3 - x_2)(x_1 - x_3)\right]\Delta$$

$$= \frac{1}{4\Delta}\left[(y_2 - y_3)(y_3 - y_1) + (x_3 - x_2)(x_1 - x_3)\right] \tag{21.48a}$$

Similarly,

$$C_{11}^{(e)} = \frac{1}{4\Delta}\left[(y_2 - y_3)^2 + (x_3 - x_2)^2\right] \tag{21.48b}$$

$$C_{13}^{(e)} = \frac{1}{4\Delta}\left[(y_2 - y_3)(y_1 - y_2) + (x_3 - x_2)(x_2 - x_1)\right] \tag{21.48c}$$

$$C_{22}^{(e)} = \frac{1}{4\Delta}\left[(y_3 - y_1)^2 + (x_1 - x_3)^2\right] \tag{21.48d}$$

$$C_{23}^{(e)} = \frac{1}{4\Delta}\left[(y_3 - y_1)(y_1 - y_2) + (x_1 - x_3)(x_2 - x_1)\right] \tag{21.48e}$$

$$C_{33}^{(e)} = \frac{1}{4\Delta}\left[(y_1 - y_2)^2 + (x_2 - x_1)^2\right] \tag{21.48f}$$

and

$$C_{21}^{(e)} = C_{12}^{(e)}, \quad C_{31}^{(e)} = C_{13}^{(e)}, \quad C_{32}^{(e)} = C_{23}^{(e)} \tag{21.48g}$$

These calculations would be easier, if the following notations are used, i.e.

and

$$\left.\begin{aligned} P_1 = (y_2 - y_3), \quad P_2 = (y_3 - y_1), \quad P_3(y_1 - y_2); \\ Q_1 = (x_3 - x_2), \quad Q_2 = (x_1 - x_3), \quad Q_3 = (x_2 - x_1) \end{aligned}\right\} \tag{21.49a}$$

By using P_i and Q_i (i = 1, 2, 3 the local node numbers), each term of the element coefficient matrix is obtained as

$$C_{ij}^{(e)} = \frac{1}{4\Delta}(P_i P_j + Q_i Q_j) \tag{21.49b}$$

where $$\Delta = \frac{1}{2}(P_2Q_3 - P_3Q_2) \tag{21.49c}$$

From Eqs. (21.49a), it should be noted that

$$P_1 + P_2 + P_3 = 0 = Q_1 + Q_2 + Q_3 \tag{21.50a}$$

$$\therefore \quad \sum_{i=1}^{3} C_{ij}^{(e)} = 0 = \sum_{j=1}^{3} C_{ij}^{(e)} \tag{21.50b}$$

This serves as a check for these calculations.

21.7.4 Assembling All Element Equations

Having considered a typical element so far, we next move on to take into account all such elements in the solution region.

Hence the energy associated with the collection of all the elements in the solution region comes out to be

$$W = \sum_{e=1}^{N} W_e = \sum_{e=1}^{N} \frac{1}{2} \varepsilon [V]^T [C][V] \tag{21.51}$$

where

$$[V] = \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_n \end{bmatrix} \tag{21.52}$$

n being the number of nodes, N the number of elements, and $[C]$ = the global (or the overall) coefficient matrix, being the collection of the individual coefficient matrices.

The main problem here is the derivation of $[C]$ from $[C^{(e)}]$. This process is best explained by considering an actual example. So we consider a finite element mesh (a very simple one) as shown in Figure 21.19, consisting of three elements which are interconnected.

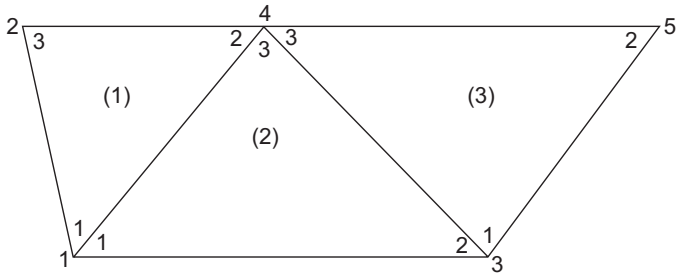


Figure 21.19 A collection of three elements.

There are three elements and five nodes. The numbering of the nodes 1–2–3–4–5 (as shown on the outside of the elements) is called the 'global' numbering. The numbering *i–j-k*, called the 'local' numbering (shown inside each element) corresponds to the sequence 1–2–3 for each of the

elements. Note that in Figure 21.19, the global numbering 3–5–4 of the element (3) corresponds to the local numbering 1–2–3 respectively of that element. It to be noted that the local numbering for each element must be the counterclockwise sense whilst there is no such restriction on the global numbering, and instead of 3–5–4 it is possible to choose 5–4–3 or 3–5–4 to correspond with the local numbering of 1–2–3. It would be seen that whatever the sense of the global numbering sequence, the global coefficient matrix [*C*] finally obtained, would be the same in all cases. Since there are 5 nodes in our present problem, the (*C*) matrix will be a 5×5 matrix of the form given as below:

$$[C] = \begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} \\ C_{21} & C_{22} & C_{23} & C_{24} & C_{25} \\ C_{31} & C_{32} & C_{33} & C_{34} & C_{35} \\ C_{41} & C_{42} & C_{43} & C_{44} & C_{45} \\ C_{51} & C_{52} & C_{53} & C_{54} & C_{55} \end{bmatrix} \tag{21.53}$$

As in the case of the element coefficient matrix, C_{ij} represents the coupling between the nodes i and j.

C_{ij} is evaluated by using the fact that 'the potential disribution across the inter-element boundaries must be continuous'.

The contribution to the position i–j in the [*C*] matrix comes from all the elements which contain the nodes i and j. As an example, to evaluate C_{11}, from Figure 21.19, the global node 1 belongs to the elements (1) and (2) and it also happens to be the local node 1 in both the elements. Hence

$$C_{11} = C_{11}^{(1)} + C_{11}^{(2)} \tag{21.54a}$$

For C_{22} it belongs only to the element (1) and is same as the local node 3, and hence

$$C_{22} = C_{33}^{(1)} \tag{21.54b}$$

Similarly for C_{55}, it belongs to the element (3) only, and is same as the local node 2, and so

$$C_{55} = C_{22}^{(3)} \tag{21.54c}$$

Considering the global node 3, it belongs to elements (2) and (3) and is the local node 2 of element (2) and the local node 1 of the element (3). Thus,

$$C_{33} = C_{22}^{(2)} + C_{11}^{(3)} \tag{21.54d}$$

The global node 4 belongs to all the three elements (1), (2) and (3); and is the local node 2 of the element (1), the local node 3 of the element (2) as well as the local node 3 of the element (3); and hence

$$C_{44} = C_{22}^{(1)} + C_{(33)}^{(2)} + C_{33}^{(3)} \tag{21.54e}$$

There is no direct link (or coupling) between

(a) the global nodes 2 and 3, and hence $C_{23} = 0 = C_{32}$
(b) the global nodes 1 and 5, and hence $C_{15} = 0 = C_{51}$
(c) the global nodes 2 and 5, and so $C_{25} = 0 = C_{52}$. (21.55a)

For C_{12} the global link 12 is same as the local link 13 of the element (1)

$$\therefore \quad C_{12} = C_{13}^{(1)} \tag{21.56a}$$

and similarly for C_{21}, the global link 21 is same as the local link 31 of the element (1)

$$\therefore \quad C_{21} = C_{31}^{(1)} \tag{21.56b}$$

For C_{13}, the global link 13 coincides with the local link 12 of the element (2)

$$\therefore \quad C_{13} = C_{12}^{(2)} \tag{21.56c}$$

For C_{31}, the same argument holds except that the linking direction is reversed.

$$\therefore \quad C_{31} = C_{21}^{(2)} \tag{21.56d}$$

For C_{14}, the global link 14 coincides with the local link 12 of the element (1) and the local link 13 of the element (2). Hence

$$\therefore \quad C_{14} = C_{12}^{(1)} + C_{13}^{(2)} \tag{21.56e}$$

For C_{41}, only the linking direction is reversed in the above Eq. (21.56e)

$$\therefore \quad C_{41} = C_{21}^{(1)} + C_{31}^{(2)} \tag{21.56f}$$

Similar arguments for the remaining links will give

$$C_{34} = C_{(23)}^{(2)} + C_{13}^{(3)} \tag{21.56g}$$

$$C_{43} = C_{(32)}^{(2)} + C_{31}^{(3)} \tag{21.56h}$$

$$C_{35} = C_{12}^{(3)} \quad \text{and} \quad C_{53} = C_{21}^{(3)} \tag{21.56k}$$

$$C_{54} = C_{23}^{(3)} \quad \text{and} \quad C_{45} = C_{32}^{(3)} \tag{21.56l}$$

$$C_{42} = C_{23}^{(1)} \quad \text{and} \quad C_{24} = C_{32}^{(1)} \tag{21.56m}$$

Having thus obtained all the terms of the global coefficient matrix, it can now be written as:

$$[C] = \begin{bmatrix} C_{11}^{(1)} + C_{11}^{(2)} & C_{13}^{(1)} & C_{12}^{(2)} & C_{12}^{(1)} + C_{13}^{(2)} & 0 \\ C_{31}^{(1)} & C_{33}^{(1)} & 0 & C_{32}^{(1)} & 0 \\ C_{21}^{(2)} & 0 & C_{22}^{(2)} + C_{11}^{(3)} & C_{23}^{(2)} + C_{13}^{(3)} & C_{12}^{(3)} \\ C_{21}^{(1)} + C_{31}^{(2)} & C_{23}^{(1)} & C_{32}^{(2)} + C_{31}^{(3)} & C_{22}^{(1)} + C_{33}^{(2)} + C_{33}^{(3)} & C_{32}^{(3)} \\ 0 & 0 & C_{21}^{(3)} & C_{23}^{(3)} & C_{22}^{(3)} \end{bmatrix} \tag{21.57}$$

In the global coefficient matrix [C], there are 27 terms, i.e. 9 for each of the 3 elements, and the element coefficient matrices overlap at those nodes which are shared by the elements.

The properties of the global coefficient matrix are:

1. It is symmetric, i.e. $C_{ij} = C_{ji}$ like the element coefficient matrix.
2. $[C]$ is sparse and banded.
   The reason for sparseness is that $C_{ij} = 0$ for those values of i and j where no coupling exists between the two nodes.
3. It is singular.

21.7.5 Solving the Resulting Global Equations

From the discussion of the variational basis of FEM in Chapter 5 and the earlier part of this chapter (Chapter 5, Sections 5.6.1–5.6.5, and Chapter 21, Section 21.2), we know that the Laplace's equation (or to be more general the Poisson's equation as the Laplace's equation is a degenerate case of Poisson), is satisfied when the total energy of the problem domain is minimized. This implies that the partial derivatives of W (= the total energy) with respect to each nodal value of the potential (V_i, i = 1, 2,..., n) are zero, i.e.

$$\frac{\partial W}{\partial V_1} = \frac{\partial W}{\partial V_2} = ... = \frac{\partial W}{\partial V_n} = 0$$

or

$$\frac{\partial W}{\partial V_k} = 0, \qquad k = 1, 2, ..., n \tag{21.58}$$

As an example, to evaluate $\dfrac{\partial W}{\partial V_1}$ for the three element mesh of Figure 21.19, W is given in Eq. (21.51) where N (the number of elements) = 3 and n the number of nodes) is 5, i.e.

$$W = \sum_{e=1}^{N} \frac{1}{2} \varepsilon [V]^T [C][V] \tag{21.51}$$

Since the partial differentiation is with respect to V_1, only these terms of C which contain the node 1 effect will give non-zero values, the rest of the derivatives will be zeroes. i.e.

$$\frac{\partial W}{\partial V_1} = V_1C_{11} + V_2C_{12} + V_3C_{13} + V_4C_{14} + V_5C_{15}$$
$$+ V_1C_{11} + V_2C_{21} + V_3C_{31} + V_4C_{41} + V_5C_{51} = 0$$

Since $C_{ij} = C_{ji} (j \neq i)$, the above equation simplifies to

$$V_1C_{11} + V_2C_{12} + V_3C_{13} + V_4C_{14} + V_5C_{15} = 0 \tag{21.59}$$

$\therefore$ In general $$\frac{\partial W}{\partial V_k} = 0,$$

gives

$$\sum_{i=1}^{n} V_i C_{ik} = 0 \tag{21.60}$$

n being the number of nodes in the mesh.

This equation [i.e. (21.60)] can be written for all the nodes $k = 1, 2,..., n$ ($n = 5$ in the present case) and thus a set of simultaneous equations are obtained from which the solution of

$$[V]^T = [V_1, V_2, ..., V_n]$$

can be found. This can be done by any of the two methods briefly described below.

1. Iteration Method

This method is similar to that used in F.D.M. Let the node 1 (of Figure 21.19) be a free node. The potential at the node 1 can be obtained from Eq. (21.59) by writing it as

$$V_1 = \frac{1}{C_{11}} \sum_{i=2}^{5} V_i C_{1i} \tag{21.61}$$

In general, if the kth node is made free, then

$$V_k = \frac{1}{C_{kk}} \sum_{i=1, i \neq k}^{n} V_i C_{ik} \tag{21.62}$$

This equation is applied iteratively to all the free nodes in the mesh which of course contains n nodes. But if the node k is not directly linked with the node i, then $C_{ki} = 0$. So only these nodes, which are directly linked with the node k, would contribute to V_k in the above Eq. (21.62).

Thus the potential of the node k, i.e. V_k can be determined by using Eq. (21.62), provided the potentials at the nodes, which are connected to the node k, are known. So the iteration process is started by setting the potentials at the free nodes to 'zero' or to the average potential which is

$$V_{\text{av}} = \frac{1}{2}(V_{\min} + V_{\max}) \tag{21.63}$$

where $V_{\max}$ and $V_{\min}$ are the maximum and minimum values of the prescribed (or defined) potentials at the fixed nodes. Using these as the initial values, the potentials at the free nodes are calculated from Eq. (21.62). When the first iteration is complete, the calculated new value of potentials at all the free nodes become the starting values for the second iteration. This process is repeated till the change between the subsequent iterations becomes negligible, i.e. for the pth and $(p + 1)$th iteration.

$$\left| V_k^{p+1} - V_k^p \right| < \varepsilon, \qquad \text{for } k = 1, 2, 3, ..., n$$

where ε can be made as small as possible (e.g. $\underline{\varepsilon} \sim .0001$ or 1×10^{-7}, depending on the accuracy requirements).

2. Band Matrix Method

In this method, all the free nodes are numbered first and the fixed nodes last, so that Eq. (21.5l) can be rewritten as

$$W = \frac{1}{2} \varepsilon [v_f \quad v_p] \begin{bmatrix} C_{ff} & C_{fp} \\ C_{pf} & C_{pp} \end{bmatrix} \begin{bmatrix} V_f \\ V_p \end{bmatrix} \tag{21.64}$$

where the subscripts f and p refer to nodes with free (floating) and prescribed (a fixed) potentials. As V_p is constant (since it consists of known fixed values), the differentiations will be only with respect to V_f. Thus applying Eqs. (21.58) to (21.64), the result is

$$C_{ff}V_p + C_{fp}V_p = 0 \tag{21.65}$$

or, as we are dealing with matrices, this is

$$[C_{ff}][V_f] = -[C_{fp}][V_p] \tag{21.66}$$

This equation can be expressed in the form

$$[A]\,[V] = [B] \tag{21.67a}$$

or

$$[V] = [A]^{-1}\,[B] \tag{21.67b}$$

where $[V] = [V_f]$, $[A] = [C_{ff}]$, and $[B] = -[C_{fp}]\,[V_p]$

$[A]$ is, in general, non-singular, and hence the potential at the free nodes can be evaluated by using Eq. (21.66). In Eq. (21.67a), $[V]$ can be solved for by Gaussian elimination. $[V]$ in the equation (21.67b) can be solved by matrix inversion, provided the matrix to be inverted is not large.

21.7.6 Areal Coordinate System and 2D Linear Triangular Elements

21.7.6.1 Areal coordinate system

We will first define this coordinate system.

Let ABC be a triangle and P be any point (either inside or outside the triangle). Then the areal coordinates of the point P with reference to the triangle ABC are defined as the ratios

$$\frac{\Delta PBC}{\Delta ABC}, \quad \frac{\Delta PCA}{\Delta BCA}, \quad \frac{\Delta PAB}{\Delta CAB} \tag{21.68}$$

Let these three ratios be denoted by X, Y, Z respectively. It is obvious that if P lies within the triangle ABC, then

$$X + Y + Z = 1 \tag{21.69}$$

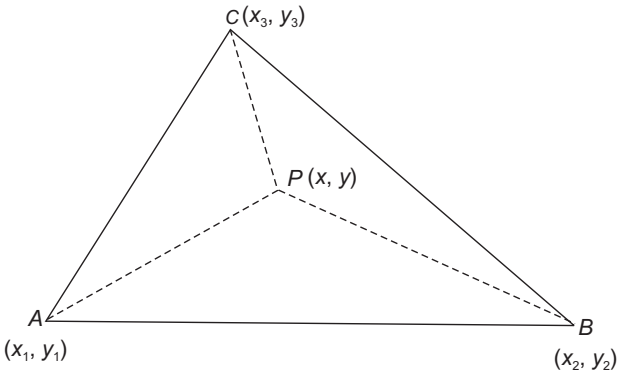


Figure 21.20 Triangle *ABC* to define the areal coordinate system.

But the location of P is not a necessary constraint for Eq. (21.69), and this relationship holds wherever the point P may be in the plane of the triangle ABC (either inside or outside of ΔABC). For this the signs of the areas of the triangles have to be accounted for. If P and A are on the same side of line BC, then both the areas PBC and ABC have the same sign and the ratio X will be +ve. But if P and A are on the opposite side of BC, then the ratio of the areas will be –ve.

Thus X will be +ve or –ve according as P and A are on the same or opposite sides of BC.

Similarly Y will be +ve or –ve according as P and B are on the same or opposite sides of CA.

And also Z will be +ve or –ve according as P and C are on the same or opposite sides of AB.

It should be noted that the three denominators for the three coordinates X, Y, Z are the same both in magnitude and sign, e.g.

$$\Delta ABC = \Delta BCA = \Delta CAB,$$

the cylic order of the letters being the same in all. (anticlockwise rotation, referring to Figure 21.20). Hence,

$$\Delta PBC + \Delta PCA + \Delta PAB = \Delta ABC$$

whatever be the position of the point P.

$$\therefore \qquad X + Y + Z = 1 \tag{21.69a}$$

Next, without going into rigorous mathematical derivations, we will now state some important and interesting results pertaining to this coordinate system.

1. If (x_1, y_1), (x_2, y_2), (x_3, y_3) be the Cartesian coordinates referred to any axes (rectangular or oblique) in its plane of the vertices of the triangle ABC, and (x, y) the Cartesian coordinates of the point P in its plane, then

$$\begin{aligned} x &= x_1X + x_2Y + x_3Z \\ y &= y_1X + y_2Y + Y_3Z \end{aligned} \tag{21.70}$$

   where X, Y, Z are the areal coordinates of the point P referred to the triangle ABC.
2. The general equation of a straight line in Cartesian coordinates is

$$Ax + By + C = 0 \tag{21.71a}$$

   which when transformed to areal coordinates, it becomes of the form

$$LX + MY + NZ = 0 \tag{21.71b}$$

3. The equation of a line at infinity is often written as

$$X + Y + Z = 0 \tag{21.72a}$$

   though in rigorous form, it is

$$X + Y + Z = \lim_{\varepsilon \to 0} \varepsilon (lX + mY + nZ) \tag{21.72b}$$

4. By using the relation $X + Y + Z = 1$, the algebraic equation of any curve in areal coordinates can be made homogeneous in X, Y, Z.

Note: Areal coordinates are only a particular case of a general system of 'homogeneous coordinates'.

In the general homogeneous coordinates, (X, Y, Z) the coordinates of the point P whose Cartesian coordinates are (x, y) would have the relationship

$$\left.\begin{aligned} x &= \lambda_1 X + \mu_1 Y + \nu_1 Z \\ y &= \lambda_2 X + \mu_2 Y + \nu_2 Z \end{aligned}\right\} \tag{21.73}$$

where X, Y, Z are related by the equation

$$1 = \alpha X + \beta Y + \gamma Z \tag{21.74}$$

(in areal coordinates, $\alpha = \beta = \gamma = 1$)

Now
$$X = \frac{1}{\alpha}\frac{\Delta PBC}{\Delta ABC}, \quad Y = \frac{1}{\beta}\frac{\Delta PCA}{\Delta BCA}, \quad Z = \frac{1}{\gamma}\frac{\Delta PAB}{\Delta CAB} \tag{21.75}$$

The detailed discussion and derivation of both the coordinate systems would be found in E.H. Askwith's *Analytical Geometry of the Conic Sections.*

21.7.6.2 Shape functions and 2D linear triangular element

In two-dimensional F.E.M., triangle has been used as the basic element shape and a suitable trial function based on this element has to be obtained. Consider the triangle of Figure 21.21, where

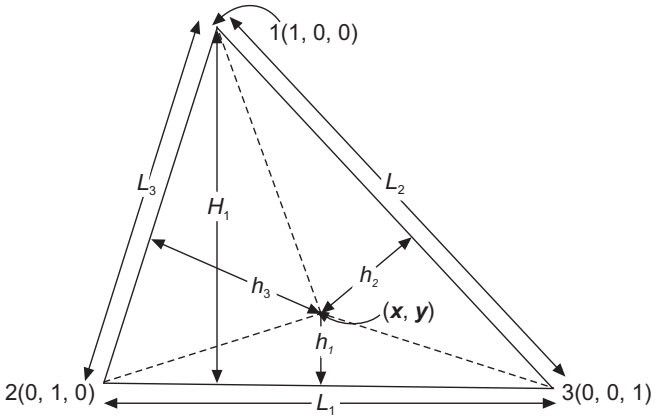


Figure 21.21 Triangular element with its shape function dimensions.

$$N_i = \frac{h_i}{H_i}, \quad i = 1, 2, 3 \tag{21.76}$$

and (x, y) corresponds to (N_1, N_2, N_3). (21.77)

The natural first order interpolation to be used within the triangle is

$$\phi(x, y) = N_1\phi_1 + N_2\phi_2 + N_3\phi_3 \tag{21.78}$$

The relationship which shows the interdependence of the three N's can be determined by adding up the areas of the three triangles formed by joining the vertices of the triangular element to the point (x, y) as shown in Figure 21.21, i.e.

$$\frac{1}{2}h_1L_1 + \frac{1}{2}h_2L_2 + \frac{1}{2}h_3L_3 = A \tag{21.79}$$

where A is the total area of the triangle.

Replacing h_1, h_2 and h_3 the trial functions given in Eq. (21.76), we obtain

$$\frac{1}{2}N_1H_1L_1 + \frac{1}{2}N_2H_2L_2 + \frac{1}{2}N_3H_3L_3 = A \tag{21.80}$$

But for every $i = 1, 2, 3$

$$A = \frac{1}{2}H_1L_1 = \frac{1}{2}H_2L_2 + \frac{1}{2}H_3L_3 \tag{21.81}$$

Combining Eqs. (21.80) and (21.81), we get

$$N_1 + N_2 + N_3 = 1 \tag{21.82}$$

Two more relations between the triangular coordinates can be obtained in terms of the Cartesian coordinates. Since (x, y) is of order one, these coordinates may be fitted exactly, with no inherent approximation, by the interpolation of (21.82), so that

$$x_1N_1 + x_2N_2 + x_3N_3 = x \tag{21.83a}$$

and

$$y_1N_1 + y_2N_2 + y_3N_3 = y \tag{21.83b}$$

Note:

1. Compare the similarity between the above relations and the equations of the areal coordinates discussed in Section 21.7.6.1.
2. Areal coordinate system is a three-axis, non-orthogonal, coordinate system for representation of two-dimensional regions so that all the relationships can be expressed in homogeneous form.

21.8 A SHORT NOTE ON FINITE-DIFFERENCE TIME-DOMAIN METHOD (F-D T-D)

21.8.1 Introduction

The interaction of electromagnetic waves with arbitrarily shaped objects cannot be modelled so easily nor accurately. Antenna apertures, cavity resonators, material inhomogeneities, etc. are some of the examples of engineering interest. The modal transmission in such structures cannot be modelled numerically so easily as to obtain correct information regarding magnitude and phase. The methods for modelling such structures have been based on frequency domain integral equation approach. The method of moments (MOM) and the finite element method (FEM) are of this category. But these methods may require large matrices to be inverted. Hence for such structures the F-D T-D method has certain advantages.

The F-D T-D method gives a direct solution of Maxwell's time-dependent curl equations (i.e. Faraday's law and modified Ampere's law equations). It employs simple second-order central difference approximation for space and time derivatives of the electric and the magnetic fields (Though for time-derivatives, sometimes forward-difference approximation is preferred.) The F-D-T-D is a time marching process which simulates the continuous actual waves by sampled-data numerical analogues stored in the computer.

To illustrate the technique of the FD-TD method, a simple problem has been considered here. As an example, a simple first order partial differential equation is solved. The equation has one space variable and a time variable, i.e.

The potential function ϕ is assumed to be a function of z-coordinate (space) and time t.

$$\therefore \qquad \phi = \phi(z, t) \tag{21.84}$$

The equation under consideration is

$$\frac{\partial \phi}{\partial z} + \frac{1}{\mathrm{c}} \frac{\partial \phi}{\partial t} = 0 \tag{21.85}$$

with the initial condition

$$\phi(z, t = 0) = F(z) \qquad \text{(say)} \tag{21.86}$$

(This equation is sometimes called the advection equation.)

It should be noted that the analytical solution of this equation is

$$\phi(z, t) = F(z - \mathrm{c}t) \tag{21.87}$$

which is a wave travelling in the $+z$-direction.

A point of further interest is that Eq. (21.87) is also a solution of the one-dimensional wave equation where equation is of the form

$$\frac{\partial^2 \phi}{\partial z^2} - \frac{1}{\mathrm{c}^2} \frac{\partial^2 \phi}{\partial t^2} = 0 \tag{21.88}$$

Both the wave equation (i.e. Eq. (21.88) and the advection equation (i.e. Eq. (21.85), belong to the family of differential equations known as 'hyperbolic equations'. The scalar diffusion equation (or heat conduction equation) is a 'parabolic equation' and Laplace's and Poisson's equations are called 'elliptic equations'. The vector eddy current equation, with harmonic time variation (i.e. $e^{j\omega t}$) can be reduced to 'elliptic equation with complex variables'.

A three-dimensional view of ϕ variation can be obtained pictorially with z and t as two coordinate axes and ϕ along the third axis orthogonal to z and t as shown in Figure 21.22.

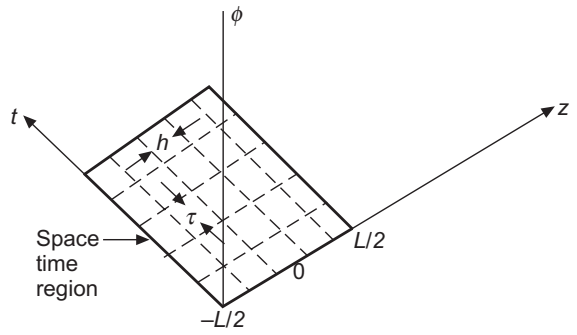


Figure 21.22 Grid system used for evaluation of ϕ of Eq. (21.85).

Since the problem is to be solved numerically, the space dimension (say L) in which the wave is propagating, has been sub-divided into N sections, so that

$$h = \frac{L}{N} \tag{21.89}$$

It the velocity of propagation is c, then let τ be the time-period required to transverse the length h of the grid, so that

$$h = \mathrm{c}\tau \tag{21.90}$$

At this stage, the numerical stability is not be discussed; and assuming 'periodic boundary condition', i.e. when the wave reaches the boundary $z = L/2$, it reappears at the same time at $z = -L/2$, and thus continues to propagate in the region $-L/2 \le z \le +L/2$.

Next, Eq. (21.85) is converted to F.D. form. This can be done by Taylor series expansion or by line integration, as shown in Chapter 5, Section 5.5.2). Hence without going through the intermediate steps again, the final result is written down directly here.

The time derivative term is approximated by using forward difference,

$$\frac{\partial \phi}{\partial t} = \frac{f(z_i, t_n + \tau) - f(z_{i,}\, t_n)}{\tau} \tag{21.91}$$

where

$$z_i = \left(i - \frac{1}{2}\right)h - \frac{L}{2}, \qquad t_n = (n-1)\tau \tag{21.91a}$$

referring to Figure 21.22.

The space derivative is approximated by using central differences,

$$\frac{\partial \phi}{\partial z} = \frac{\phi(z_i + h, t_n) - \phi(z_i - h, t_n)}{2h} \tag{21.92}$$

Substituting from Eq. (21.91) and (21.92), Eq. (21.85) reduces to:

$$\frac{\phi(z_i + h, t_n) - \phi(z_i - h, t_n)}{2h} + \frac{1}{\mathrm{c}} \frac{\phi(z_i, t_n + \tau) - \phi(z_i, t_n)}{\tau} = 0 \tag{21.93}$$

Of the four terms in the above equation, three terms are at the time instant t_n and one term is at the instant $t_n + t$. Hence rewriting the equation,

$$\phi(z_i, t_n + \tau) = \phi(z_i, t_n) - \frac{\mathrm{c}\tau}{2h}\left[\phi(z_i + h, t_n) - \phi(z_i - h, t_n)\right] \tag{21.94}$$

The finite difference method based on this equation is the F-D T-D method. The evaluation of these terms in the time-steps is the 'Time-marching process'. The initial values are all known at $t_0 = 0$, and the boundary conditions would give the values at the boundaries.

This is the underlying principle of the F-D T-D method, and serves as an introduction to this method.

22 Modern Topics and Applications

22.1 INTRODUCTION

In our study of electromagnetism, we have so far discussed both the magnetic diffusion (i.e. eddy currents) and most of the aspects of electromagnetic waves (i.e. propagation, guidance and radiation). Apart from the applications discussed till now, there are other areas of applications of electromagnetism. Some of these applications are microwaves used in divese areas ranging from satellite communication, EM interference and compatibility, remote seusing, heating, radar, radio astronomy and so on, lasers, fibre optics, etc on the high frequency side, and electrical machines and allied power generating equipment on the low frequency side. We shall now briefly describe and discuss some of the new applications since the old applications have been comprehensively discussed in specialized textbooks elsewhere.

We shall start with some of the important high frequency applications first and then go on to describe briefly the new power frequency applications as well.

22.2 MICROWAVES

Microwaves are electromagnetic waves within the frequency range of approximately 300 MHz to 1000 GHz.

In the field of communication, the main means for carrying thousands of channels over long distances are:

(1) Microwave links (2) Co-axial cables, and (3) Optical fibres (this being a new device, it will be discussed briefly later.

Microwaves are highly attractive for communications because of the following properties:

(a) Wide available bandwidth
(b) Directive properties of short wavelengths.

Because of these properties, microwave communication has now become more common than the radio and TV. A microwave system usually consists of a transmitter which includes mocrowave oscillator, waveguides and a transmitting antenna and receiver sub-system made up of receiving antenna, transmission line or waveguide, microwave amplifier and a receiver. The use of microwaves

has very significantly increased. Some of the examples are telecommunications, radio astronomy, land surveying (by remote sensing). Radar, terrestrial microwave links, heating, medicines and so on. We shall briefly describe qualitatively the salient aspects of some of these applications. We shall also describe some of the important low-frequency applications.

22.2.1 Microwave Heating

Microwaves are preferred over the low frequency electromagnetic waves because the microwave energy can be more easily directed, controlled and concentrated. Also various atomic and molecular resonances occur at the microwave frequencies. So this energy can be used for heating purposes (another being remote sensing). The heating property of microwave power is used in a wide variety of domestic, commercial and industrial applications. The microwave oven is one such typical example (Figure 22.1).

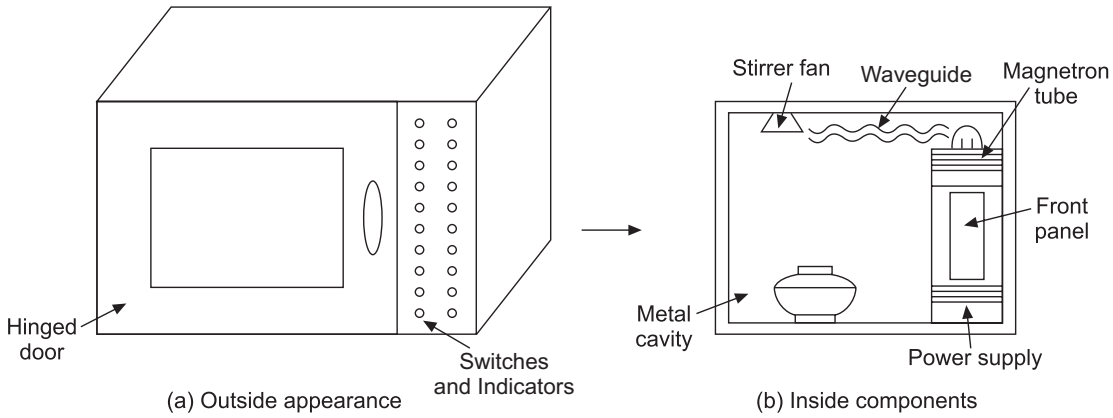


Figure 22.1 Microwave oven (a typical one).

The oven gets the microwave energy for cooking from a magnetron which is a thermionic valve capable of producing high power oscillations in the microwave region. The magnetron consists of a heater, central cathode and a concentric, radially segmented anode, all enclosed in an evacuated container, located in the gap of an external magnet. The movement of electrons is controlled by a combination of crossed electric and magnetic fields. It is also used extensively in radars. The energy from the magnetron is carried through a waveguide into the oven which is basically a resonant cavity in which the food for cooking is located. The reflections from the stationary walls and the motion of the stirring fan ensure the microwave energy to be well distributed in the cavity. The shape of the oven cavity is usually a rectangular parallelepiped, though there are some ovens with cylindrical cavity having a hemispherical top for more uniform cooking. Thus the microwave makes the cooking to be fast and evenly distributed. In modern microwave ovens, facilities are provided for grilling, convective cooking as well, in addition to the microwave cooking. These two additional facillities are done at normal frequency and so these sources (i.e. heater and fan) get their electricity at power frequency, thus bypassing the magnetron. The microwave oven operates at 2.4 GHz. Apart from this type of domestic cooking, the microwave heating capability is used in physical diathermy (in medical treatment) and in drying potato crisps, paper, cloth, etc.

22.2.2 Radar Systems

The word radar is an acronym for **RA**dio **D**etection **A**nd **R**anging. It includes any system that employs microwaves for the purpose of locating, identifying, navigating or guiding such moving objects as ships, aircraft, missiles, or artificial satellites. It is now also used for burgler alarms, garage-door openers, and police speed detectors.

The system essentially consists of a generator of electromagnetic radiation of centimetre wavelengths, the output of which is pulse-modulated at a radio frequency and fed to a movable antenna from where it is radiated as a beam. Distant objects crossing the path of the beam reflect the pulses back to the transmitter which also acts as a receiver. A cathode-ray tube (CRT) indicator displays the received signal in the correct time sequence so that the time required for a pulse to travel to the object and back can be measured. Thus the distance of the object from the transmitter can be calculated and its direction can be found from a knowledge of the direction of the antenna. This fundamental technique has now been extrapolated so that automatic guidance and navigation can also be made by computers without requiring a display system.

The radio transmitter should be of high power and the receiver has to be extremely sensitive. The receiving antenna is also highly directional. The amplitude and delay of the receiving signal are the two important parameters required for estimating the size and the distance of the observed object from the radar. The size of the object is characterized by a parameter called the "radar cross-section". This is a complex function of the reflectivity of the object and its physical shape and size. For example, metallic objects have higher radar cross-section compared with dielectric objects.

Power reflected by an object

= Power density of the incident wave × Radar cross-section (= σ)

Let

P_t = Transmitting power of the antenna

G = Directive gain of the antenna

A_e = Effective aperture of the antenna in the receiving mode $= \dfrac{G\lambda^2}{4\pi}$ (since the same antenna is used for transmission and reception)

∴ Power density of the transmitted signal (at the object),

$$W = \frac{P_t G}{4\pi r^2} \tag{22.1}$$

where r is the distance of the object from the radar.

∴ Power reflected by the object,

$$P_{\text{ref}} = P_\sigma = \frac{P_t G \sigma}{4\pi r^2} \tag{22.2}$$

∴ Power density of the reflected power at the radar,

$$W_r = \frac{P_{\text{ref}}}{4\pi r^2} = \frac{P_t G \sigma}{(4\pi r^2)^2} \tag{22.3}$$

(This assumes that the reflected power is uniformly distributed in all directions.)

∴ Power received by the radar antenna,

$$P_{\text{rec}} = W_r A_e = \frac{P_t G\sigma}{(4\pi r^2)^2} \frac{G\lambda^2}{4\pi} = \frac{P_t \sigma}{4\pi} \left(\frac{G\lambda}{4\pi r^2} \right)^2 \tag{22.4}$$

Thus P_{rec} must be greater than the minimum detectable power of the receiver, for detection of objects.

Since $r >> \lambda$, G must be quite large so as to maintain P_t within acceptable limits.

Hence radars use highly directional antennae like parabolic dishes.

The energy transporting medium from the transmitter to the antenna is usually a metal wave-guide (rectangular or circular) since the radar needs high power to be transmitted.

The radars, though originally developed for military purposes, now have many civilian applications, like weather monitoring, geological survey, air-traffic control, remote sensing and planetary explorations. In defence applications, when they talk of invisible bombers, it is meant that these planes are not visible on the radar.

In a sense, the defence requirements on radars are contradictory, because on one side the radar research aims at detecting targets with highest resolution and sensitivity whilst on the other side aircraft and ships are being developed with extremely low radar cross-section so as to make them as invisible to radar detection as possible. (Invisible planes are used for spying purposes).

22.2.3 Remote Sensing

In radio remote sensing, the microwave frequencies are used for geological survey (i.e. to probe the terrain). A radar is mounted on a moving platform which can be an aircraft or a satellite above the surface of the earth. Proper frequencies have to be selected for effective imaging of a region since the reflectivity of earth's surface is a function of both location as well as frequency.

Two types of radars are used for getting high quality radio images of the earth's surface, i.e.

1. Side looking radar (SLR)
2. Synthetic aperture radar (SAR).

In a SLR, the reflection from a region illuminated by the beam of the radar gives the average reflectivity of that region. As the radar carrier moves, the beam spot on the region also moves and thereby the reflectivity profile of the region along a strip is produced. Multiple scanning of adjacent strips is done to build an image of the complete region. To obtain high resolution, large antennae have to be used, and the radar vehicle should be at low attitude for better linear resolution in the images.

In the SAR, the reflected signals for the entire flight of the radar are stored and then coherently processed to form a synthesized antenna beam which is much narrower compared to the beam of the radar antenna. The linear resolution of the SAR is proportional to the size of the antenna, i.e. the resolution of the image improves as the size of the antenna is made smaller. Also the resolution is independent of the height of the vehicle.

Numerical electromagnetic techniques are used to obtain accurate models of the earth's surface.

22.2.4 Radio Astronomy

This is the study of heavenly bodies by the reception and analysis of the 'radio frequency electromagnetic radiaton' which they emit or reflect. In general the electromagnetic radiations from

the extra-terrestrial sources are either absorbed by the earth's atmosphere or reflected away from the earth by the ionosphere. The two exceptions are: (i) the optical wavelengths which are able to penetrate the atmosphere, and (ii) the radio wavelengths in the band 1 cm to 10 metres which are too long to be absorbed by the atmosphere and too short to be absorbed by the ionosphere. These radiations which pass through the 'radio window' come from a variety of sources ranging from objects within the solar system to galaxies which are too far to be observed by optical telescopes. The origin of these radio frequency emissions can be thermal or non-thermal. Information is extracted from the apparently incoherent radio noise from the universe, by constructing maps of the sky in terms of radio emission at several different frequencies which are then compared with optical observations. Thus the radio sources and radio galaxies have been identified.

Radio telescopes are the instruments used in radio astronomy to pick up and analyse the radio frequency electromagnetic. radiations of extraterrestrial sources. The two main types of radio telescopes are:

(i) *Parabolic reflectors* (paraboloid reflector—a concave reflector, the section of which is a parabola—also called a dish antenna). These are usually steerable so that they can be pointed to any part of the sky and they reflect the incoming radiation to a small antenna at the focus of the paraboloid.

(ii) *Fixed radio interferometers.* These consist of two or more separate antennae (an array), each receiving electromagnetic radiation of radio frequencies from the same source and each joined to the same receiver.

This type of telescope has greater position finding accuracy and can distinguish a small source against an intense background better than the parabolic type. On the other hand the parabolic type is more verslatile because of its mobility.

To achieve the same or similar resolution of optical telescopes, the radio telescope, operating at radio frequencies would need to have a size a few hundred metres to few kilometres. Since building a continuous aperture type of telescope of such sizes is not a viable proposition, the radio astronomers have developed a technique known as the 'aperture synthesis technique' by using which a large aperture can be synthesized without actually building one of that magnitude. The method is based on the measurment of the spatial Fourier spectrum of the brightness distribution (and not measuring this distribution directly). This spectrum is the auto-correlation function (also called visibility function) of the incoming radiation from the sky. The Fourier inversion of the visibility function then provides the brightness distribution of the sky.

22.2.5 Satellite Communication

This has been one of the important modes of long distance communication. There are many advantages of this technique, e.g. large bandwidth, mobility, dynamic assigment of resources and so on. Communication satellites are artificial earth satellites used for relaying radio, television and telephone signals around the curved surface of the earth. 'Passive' satellites merely reflect the transmissions from their surfaces whereas the 'active' satellites are equipped to receive and retransmit signals.

The orbit of an artificial earth satellite which has a period of 24 hours is called synchronous orbit (or stationary orbit). The altitude corresponding to such an orbit is about 35,700 km, and a satellite in a circular orbit parallel to the equator at this altitude would appear to be stationary in the sky (it is called a geostationary satellite). Communication satellites in synchronous orbits are used for relaying radio signals between widely separated points on the earth.

In satellite communication, radio signals are transmitted from a transmitter on the earth (this is called the 'earth station') towards the satellite. The satellite on receiving the signal, changes its frequency and retransmits it towards the earth. The satellite generally offers point to multipoint transmission (i.e. broadcasting). Generally microwave frequencies are used for satellite communication, since microwaves propagate along straight lines like a light ray and are not bent by ionosphere as are lower frequency signals. The design of a satellite communication link takes into consideration all the facts of the electromagnetic waves which have been discussed so far. The satellite link design gives an estimate of both, i.e. (a) the power which the satellite would receive from an earth station, and (b) the power which a receiving station on the ground would get from the satellite. This design would have to take into account aspects such as (i) energy absorption by the space between the satellite and the earth, (ii) changes in wave characteristics during propagation, and (iii) design of antenna system. Since the power at the satellite is limited, the antenna must be designed for high directivity as well as high efficiency. The antenna is a crucial component of this link system.

The energy loss in the path (i.e. the path-loss) is independent of the transmitting and the receiving systems. The path-loss is inversely proportional to the square of the wavelength and hence the shorter wavelengths (i.e. higher frequencies) suffer greater path-loss. A geostationary satellite operating (say) at 6 GHz would have a path-loss of about 200 dB.

22.2.6 Electromagnetic Interference and Compatibility (EMI and EMC)

Electromagnetic Interference (EMI): It is the degradation of performance and consequent malfunctioning of electronic devices and systems due to environmental electromagnetic fields.

There are various sources of electromagnetic emission, like spark plugs, relays, electric motors and generators. The disturbing fields due to such devices are comparatively localized, though there are electronic devices whose disturbing fields are more distributed. (*Note:* By disturbing fields it is meant that these fields are more of accidental by-products of the design of the device or the system. These are called 'radiated emissions'). So it can be said that the electromagnetic environment consists of apparatuses like radio and TV broadcast stations, radar systems (both air-port as well as traffic ones), flourescent lights, car ignition systems, mobile radios, navigational aids and so on, which radiate electromagnetic energy as they operate. The results of such interference are obvious in every day life, such as 'ghosting' in television reception, high voltage power lines causing interference with personal computer operation, taxicab radio interference with police radio system. Hence electromagnetic compatiblity can be defined as follows:

Electromagnetic Compatibility (EMC): An electronic system is said to 'electromagnetically compatible' with its environment when the system is able to perform compatibly with other electronic systems and neither produces nor be susceptible to the environmental electromagnetic disturbances.

To achieve EMC, all electronic devices should be able to co-exist and operate in harmony, i.e. the device should be able to function as per the designed purpose in presence of and in spite of the others. In fact EMI is the problems caused by the unwanted voltages, currents and fields on the devices and EMC is the solution to the problem. The goal of EMC is to achieve the system compatibility which is achieved by freeing the system of EMI problems.

22.2.7 Sources and Characteristics of EMI

EMI can be classified in terms of its causes and sources, so that it will be easier to control this type of disturbance. As stated earlier, any electronic device can be the source of EMI even though the designer may not have intended so. The cause of EMI can be within the system (i.e. an intrasystem problem) or can be from outside (i.e. an intersystem problem). Some examples of both the types are listed below. Here the sources have been termed emitters and the victims susceptors.

A. Intrasystem Causes		**B. Intersystem Causes**	
Emitters	*Susceptors*	*Emitters*	*Susceptors*
Power supplies	Relay	Lightening stroke	Radio receivers
Radar transmitters	Radar relays	Computers	TV sets
Mobile radio transmitters	Mobile radio receivers	Power lines	Heart pacers
Flourescent lights	Ordnance	Radar transmitters	Aircraft navigation systems
Car ignition systems	Car radio receivers		

The sources of EMI can be natural or man-made (artifical).

22.2.8 Control of EMI

There are three ways of preventing EMI and making systems compatible.

(i) Supressing the emission at the source
(ii) Making the transmission path inefficient
(iii) Making the receiver less susceptible to emission.

The first line of defence would be to avoid fast switching of the signals. This should be done even in digital systems by switching smoothly so as to reduce the high frequency content of EMI, thus reducing the coupling of the interference. Grounding establishes an electrically conducting path between two points to connect electrical and electronic elements of a system to one another or to some reference point (designated as the ground). An ideal ground plane is a zero-potential, zero-impedance body used as reference for all signals in associated circuitry. All undesired currents can be transferred to it for elimination of their effects. 'Floating point ground' is for electrical isolation from a common ground plane. But this can cause a hazard. The single-point grounding is used for minimizing the effects of gound currents and eliminating the effects of loop currents in multiple-earthing.

Shielding is done to prevent radiated energy from entering a specific region or to confine the radiated energy in a specified region (Figure 22.2). The term 'shield' usually refers to a metallic enclosure that completely encloses a particular electronic device or sub-system through non-solid and braid, as is used on cables. The shielding effectiveness is defined by:

$$\text{SE} = 10 \log_{10} \frac{\text{incident power density}}{\text{transmitted power density}} \tag{22.5}$$

where the incident power density is the power density measured at a point before the shield has been installed, and the transmitted power density is the power density at the same point after the shield has been installed. SE can also be defined in terms of electric or magnetic field strengths, i.e.

$$\text{SE} = 20 \log_{10} \frac{E_i}{E_t} \qquad \text{for electric fields}$$

and

$$\text{SE} = 20 \log_{10} \frac{H_i}{H_t} \qquad \text{for magnetic fields}$$

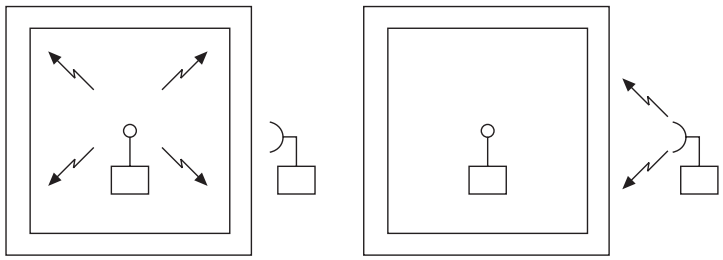


Figure 22.2 Electromagnetic shielding.

Filtering: An electrical filter is a network (either lumped or distributed) of resistors, inductors and capacitors that allows certain frequencies to pass through while blocking the other frequencies. Filters usually reduce the levels of conducted interference substantially. The most important characteristic of a filter is its insertion loss which is a function of frequency.

$$\text{IL} = 20 \log_{10} (V_1 / V_2)$$

where

V_1 = output voltage of a signal source, with the filter in the circuit
V_2 = output voltage of the signal source, without the filter in the circuit.

22.3 TELEPHONE NETWORKS

In telephone networks, all transmission is through a pair of wires. All subscriber loops in the telephone networks are made with a single pair of wires, using bidirectional transmission (i.e. the users at both ends of the line talk simultaneaously, thereby superimposing their signals on the wire pair).

For long distance transmission (i.e. between two switching offices) it is better to use two unidirectional transmissions on different pairs of wires. The reason for this requirement is that the long distance transmission requires amplifiers and these are unidirectional. Hence the long distance transmission becomes a four-wire system (Figure 22.3). So for long distance connection, at some point there has to be a conversion from two-wire to four-wire system and vice-versa. This conversion device is called a 'hybrid'.

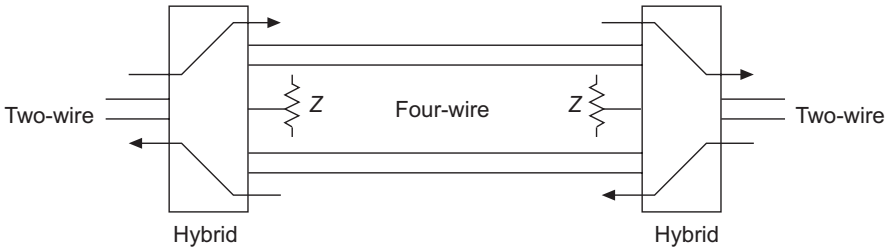


Figure 22.3 Two-wire to four-wire conncetions in telephone lines using hybrids.

In earlier days, the hybrid circuits were connected through interconnected transformers. But nowadays, the electronic hybrids need only the impedance matching for perfect isolation between oppositely travelling signals. When there is an impedance mismatch on the lines (due to the variable impedance of the two-wire line in switching environment), there would be reflections on the line causing an echo. The echoes are minimized by adaptive signal processing techniques.

22.4 LASERS

Laser is an abbreviation of the words **L**ight **A**mplification by **S**timulated **E**mission of **R**adiation. It is essentially an optical maser (see below for maser). The laser produces a powerful, highly directional, monochromatic, and coherent beam of light. Its working principle is essentially same as that of a maser, except that the 'active medium', in the case, consists of or is contained in an optically transparent cylinder with a reflecting surface at one end and a partially refltecting surface at the other. The stimulated waves pass up and down the cylinder repeatedly, some of them emerging as light through the partially reflecting end.

In a 'ruby laser', the chromium atoms of a cylindrical shaped ruby crystal are optically pumped to an excited state by a flash lamp and then it emits pulses of highly coherent light (population inversion). There are lasers which have been constructed by using a mixture of inert gases (e.g. helium and neon) to produce a continuous beam. These is also another type of laser consisting of a cube of specially treated gallium arseuide which can emit infra-red radiation when a current is passed through it. Lasers have been used in eye surgery, holography and metal cutting, etc.

Maser: This word is an acronym of **M**icrowave **A**mplification by **S**timulated **E**mission of **R**adiation. This is a class of amplifiers and oscillators which make use of the internal energy of atoms and molecules in order to obtain low-noise level amplification and microwave oscillations of precisely determined frequencies. The basic principle of operation of these devices is the stimulated emission which is the emission by an atom in an excited quantam state of a photon, as the result of the impact of a photon from outside of exactly equal energy. Then the stimulated photon or wave is augmented by the one emitted by the excited atom.

A maser consists of an active medium which can be either in the gaseous or solid state, in which most of the atoms can be optically pumped to an excited state which is done by subjecting the system to electromagnetic radiation of different frequencies to that of the stimulating frequency.

The active medium is enclosed in a resonant cavity so that a wave is built up modally (i.e. with only one mode of oscillation) thereby giving a single output frequency. Masers when made to operate at optical frequencies, are referred to as lasers. (*Note:* 'Population inversion': In a laser, when a large proportion of the emitting ions have been raised to an excited energy level by the process of optical pumping which means that the energy has been introduced in the system by an external light source, then the state of population inversion exists in it (the laser). This is an essential step in the process of stimulated emission.

It should be noted that the laser is essentially an amplifier and not a light generating device. But, since any amplifier can be converted into an oscillator by feedback with a properly selected frequency, a laser can also be converted into a source of coherent light. The crucial component of the laser is a frequency feedback device. This feedback is done by creating a resonant cavity around the active (i.e. the amplifying) medium. Figure 22.4 shows the basic elements of a laser.

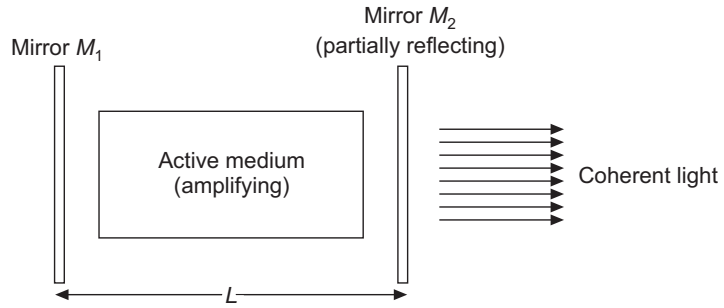


Figure 22.4 Basic elements of a laser.

Only certain frequencies of the light would get sustained amplification as a result of repeated reflection between the two mirrors M_1 and M_2. For a sustained oscillation of that frequency (= ω), the gain (= G) due to the round trip between the two mirrors should be unity and the phase difference due to this trip would be a multiple of 2π. Hence if the reflection coefficients of the two mirrors are denoted by R_1 and R_2 respectively, then

$$R_1R_2 \exp(2GL) = 1 \tag{22.6}$$

and
$$2\beta L = 2m\pi, \qquad m \text{ being an integer} \tag{22.6a}$$

where

β = phase constant of the light wave of frequency ω

$= \dfrac{\omega n}{\mathrm{c}}$, where n = refractive index of the active medium.

$\therefore$
$$\omega = \frac{m\pi \mathrm{c}}{nL} \tag{22.7}$$

The light coming out of a laser has random polarization to achieve normal polarization, the light from the laser has to be reflected from a dielectric slab at the 'Brewster angle' as mentioned in Chapter 17. Section 17.15.

22.5 OPTICAL FIBRE

Through the last century (20th), the basic media for transmitting radio frequency (RF) energy have been wire-pairs (i.e. conductors), co-axial cable, waveguides and atmosphere. Radio frequency range has been 10 kHz to 300 GHz. With increasing demands for more communication channels, a new technology has been developing and the upper limit of the radio frequency is being pushed to and beyond the infra-red and visible portion of the spectrum. This is the development and use of optical fibres which are small diameter cylindrical glass filaments which transmit EM waves at frequencies near the visible part of the spectrum. Light differs from radio in that it has both the properties of particles called photons and the properties of EM waves. 'Photonics' is the term commonly applied to optical fibres. Both particle physics and EM wave theory are necessary to understand completely the working of optical fibre systems. Particles are characterized by their energy, momentum, and they can be individually physically identified. EM waves, on the other hand are characterized by frequency, wavelength, polarization, and velocity. Basically optical fibre is a cylindrical dielectric waveguide operating at optical frequency.

Optical fibres are specially constructed glass, plastic or ceramic cylindrical waveguides to contain the propagating electromagnetic fields. The fibres operate at very high range of the frequency scale, so that it has been customary to express optical energy in terms of wavelengths (= λ). The visible spectrum ranges from 0.4×10^{-6} metres (blue) to 0.6×10^{-6} metres (red). In terms of frequency, it will be in tera hertz [= THz (10^{12})]. The most commonly used optical fibres are constructed from silica (SiO_2) which has been purified and doped in order to reduce the energy loss. The optimum performace wavelength range for SiO_2 is from 0.8×10^{-6} metres to 1.6×10^{-6} metres. An obvious advantage of operating at high frequencies is the increase in bandwidth available for transmitting information. Modulated laser bandwiths for optical fibre transmission are about 100 GHz wide, i.e. about 0.1% of the carrier frequency.

Communication systems have a wide range of power values. A laser transmitter array might be radiating 100 mW into a fibre whilst the detector at the end of the fibre might be receiving one-ten thousandth of a mW—a difference of 10^6. So the relative powers are expressed in terms of decibels.

The velocity of light (i.e. EM waves) in free space or vaccum is $v = c = 3 \times 10^8$ m/s The velocity of light in water, glass or any media other than free space is lower and is given by $v = c/n$, where n = index of refraction in the particular medium.

In every medium other than a vacuum, the velocity exhibits some degree of frequency dependence. Most RF signals have multiple frequency components usually because of modulation sidebands on a carrier. This dependence of velocity on frequency (or wavelength λ) can present significant problems. This phenomenon is called dispersion.

The indices of refraction of most materials used for different components of any optical fibre system are greater than unity (this being the refractive index of free space). In circular section optical fibres, the central section (or core) has a higher refractive index than the outer annulus (called the cladding). Usually the core index is about 1.5 and that of the cladding is slightly lower.

The velocity of the EM optical waves is less in a medium with refractive index >1 (this being the refractive index of free space). Hence, when two media of differing refractive indices meet, the optical wave is refracted or bent on the interface. Thus the wave in the second medium either speeds up or slows down depending whether its (second medium) refractive index is less than or greater than that of the first medium. The bending (or the refraction) is either towards the normal to the interface ($n_1 < n_2$) or away from the normal to the interface ($n_1 > n_2$). In an optical fibre, since the central core has higher refractive index than the outer cladding, the light which impinges on the core-cladding interface is bent back to the centre and guided along the fibre, due to total internal reflection.

The EM wave theory is used to analyse the propagation of light in the optical fibre medium. The particle theory is used for calculating the energy of the system by considering the energy of the photons. The behaviour of photons is needed for describing the laser and light-detector performances. The energy of a single photon of frequency f is

$$\left.\begin{aligned} W \text{ (joules)} &= h \times f = h \times c/\lambda \\ \text{where} \qquad h &= \text{Planck's constant} = 6.626 \times 10^{-34} \text{ J-s} \end{aligned}\right\} \tag{22.8}$$

22.5.1 An Optical System

For most of the applications the system has three components (Figure 22.5):

1. *An optical source:* There are two types of sources: (a) lasers, and (b) LEDs (Light

Emitting Diodes). Both these are electrical to optical energy converter (E/O). Lasers, at high electrical drive levels, increase their outputs through an internal oscillation. This permits high output powers and narrow source bandwidth. LEDs are semiconductors, small in size and consume little power, easier to couple to fibres and integrate, and are of low cost. Both LEDs and lasers at low drive levels emit light through spontaneous emission, which is a random process and broadband.

2. *Optical detectors:* These convert optical signals to electrical signals (O/E), so that the original signals can be recovered. These are made of semiconductor materials like Si, Ge or InGeAs. Their choice is a function of the required wavelength λ. In some detectors like avlanche photo-detector, internal gain can be achieved by using high voltages.
3. *Optical fibres* (or the media) to carry the optical signals from the source to the detector. An optical fibre is basically a cylindrical dielectric rod called core, surrounded by a dielectric annulus called cladding. The cladding has an outer jacket (i.e. buffer or coating). The electrical signals which modulate the carrier can be either analogue or digital, and are referred to as baseband input signals. The digital signals, using pulses of same level and shape (to represent the original signal) require more complicated coding, modulation/demodulation but have better noise and cost advantages. This is because of the tremendous progress in the integrated electronics uses for digital signal processing (DSP) and transmission. The electrical signals will also be part of a time-division multiplexed (TDM) bit stream, and the source might be internally or externally modulated. The source has to be carefully coupled to the fibre and might be combined with other optical signals using wavelength division multiplexing (WDM). The fibre would be probably part of a cable that contains other fibres. The fibres can be either large core multimode or small core single-mode. A long fibre would need splices at intermediate points. At the receiver end, the components are required in the reverse order to restore the electrical signal.

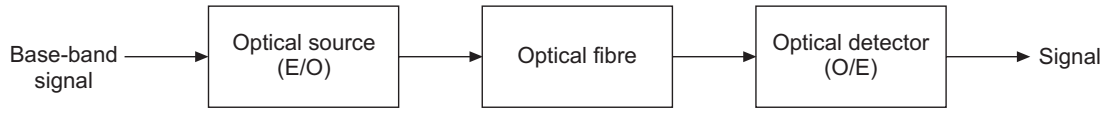


Figure 22.5 Optical fibre system.

22.5.2 Optical Fibre Characteristics

An optical fibre is made up of three concentric cylindrical sections, the core, the cladding and the buffer. The core consists of one or more thin strands of glass or plastic. The cladding is plastic or glass coating surrounding the core which may be step-index or gradient index. In the step-index core, the refreactive index is uniform but undergoes an abrupt change at the interface of the core and the cladding. The graded-index core is of the type in which the refractive index of the core varies as a function of the radial distance from the central axis of the core. The jacket or the buffer usually surrounds one or more of the cladded fibres and acts as a protection against environmental moisture or sunshine and so on (Figures 22.6 to 22.8).

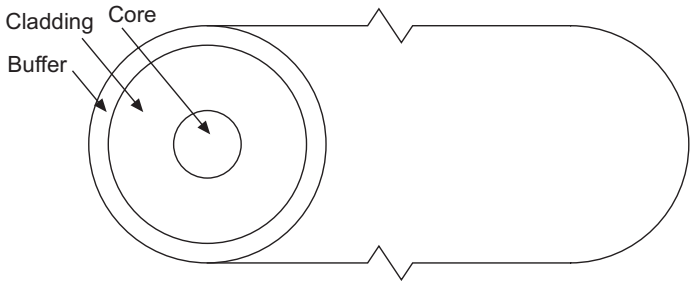


Figure 22.6 Optical fibre structure.

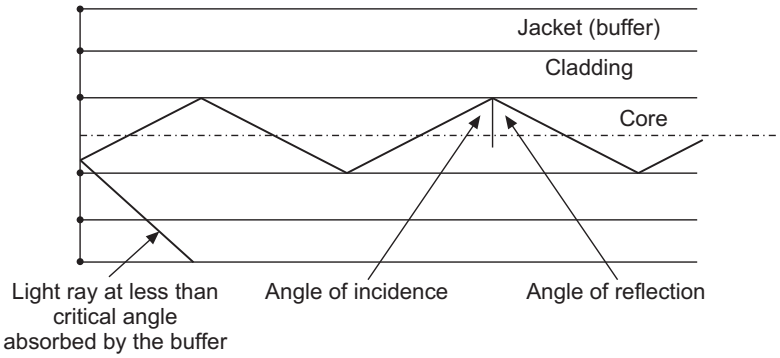
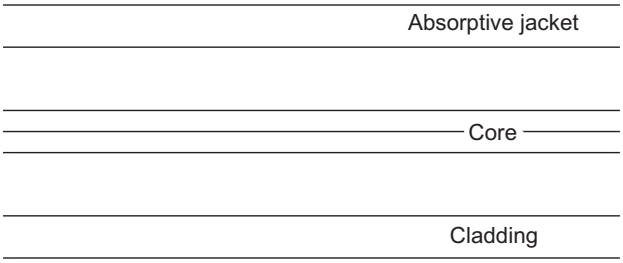
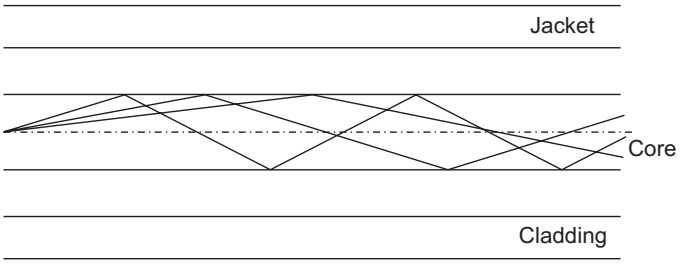


Figure 22.7 Behaviour of light ray in a fibre.

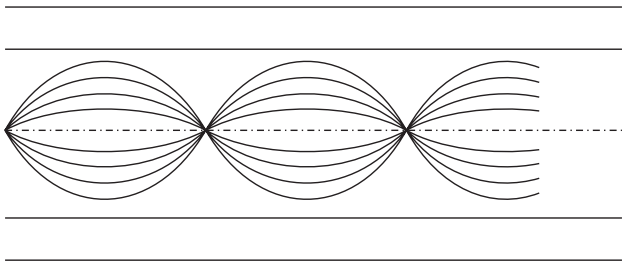


(a) Single Mode (SM)



(b) Step-index Multimode

Figure 22.8 Contd.



(c) Graded-index Multimode

Figure 22.8 Optical fibres transmission modes.

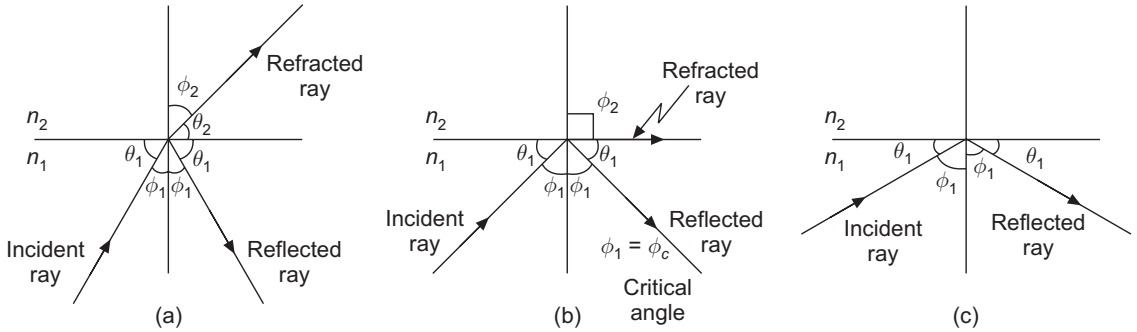


Figure 22.9 Total internal reflection, $n_1 > n_2$: (a) $\phi_1 < \phi_c$; (b) $\phi_1 = \phi_c$; (c) $\phi_1 > \phi_c$

22.5.2.1 Total internal reflection

As the light ray propagates from medium (1) to medium (2) (i.e. in the optical fibre from the core to the cladding), at the interface, the light ray (or the wavefront) will obey the Snell's law, (Figure 22.9), i.e.

$$n_1 \sin\phi_1 = n_2 \sin\phi_2 \tag{22.9}$$

(*Note:* In Chapter 17, these angles have been denoted as θs and so the ϕ mentioned here is complement of the θ in Chapter 17).

Also, since the refractive index of the core (i.e. n_1) is greater than the refractive index of the cladding (i.e. n_2) a ray incident in the core will be refracted away from the normal to the core–cladding interface [Figure 22.9(a)]. As ϕ_1 increases, there will be a critical value of ϕ_1, i.e. $\phi_1 = \phi_c$, for which $\phi_2 = 90°$. For values of $\phi_1 > \phi_c$, there will be no refracted ray and all the energy will be reflected back in the core with no leakeage in the cladding [Figures 22.9(b) and (c)].

i.e. $$\phi_1 = \text{incident angle} = \phi_c = \sin^{-1}\left(\frac{n_2}{n_1}\right) \quad \text{[Ref. Eq. (17.155)]} \tag{22.10}$$

A ray striking the interface at $\phi_1 < \phi_c$ will have a large part of its energy lost in the cladding. Since n_1 and n_2 are very close, a useful mathematical relationship can be written down as

$$\Delta = \frac{n_1^2 - n_2^2}{2n_1^2} \simeq \frac{n_1 - n_2}{n_1} \tag{22.11}$$

or $$n_2 \simeq n_1(1 - \Delta) \tag{22.11a}$$

It follows from the above that a very small Δ means that the critical angle ϕ_c is close to 90°. This is typical of silica fibres. In all plastic fibre with a large core diameter, Δ is much greater and ϕ_c is much smaller. Both these types have important applications.

For the present discussion, the 'plane of incidence' of the incident ray is as defined in Chapter 17, Section 17.13. Any linearly polarized wave of arbitrary polarization can be resolved into parallel and perpendicular (normal) components with reference to the plane of incidence. Hence the coefficients of energy reflection can be easily written down from the Fresnel's laws of reflection [Ref. Eqs. (17.153a–d)].

$$R_{\mathrm{P}}\ (\text{Parallel polarization}) = R_{\mathrm{P}} = \left| \frac{-n_2^2 \cos\phi_1 + n_1\sqrt{(n_2^2 - n_1^2 \sin^2\phi_1)}}{n_2^2 \cos\phi_1 + n_1\sqrt{(n_2^2 - n_1^2 \sin^2\phi_1)}} \right|^2 \tag{22.12a}$$

$$R_{\mathrm{N}}(\text{Normal polarization}) = R_{\mathrm{N}} = \left| \frac{n_1 \cos\phi_1 - \sqrt{(n_2^2 - n_1^2 \sin^2\phi_1)}}{n_1 \cos\phi_1 + \sqrt{(n_2^2 - n_1^2 \sin^2\phi_1)}} \right|^2$$

It should be noted that the glass-to-air interface has a critical angle of 41.8° and the air-to-glass interface has no critical angle.

With parallel polarization for both of them, there is Brewster angle for which there is ZERO reflection, a condition obtained by making the numerator of R_{P} equal to zero, so that

$$\phi_{iB} = \text{Brewster angle} = \tan^{-1}\frac{n_2}{n_1} \tag{22.13}$$

22.5.2.2 Numerical aperture

This is one of the most important parameters of an optical fibre (or any optical device which accepts and uses light energy). The numerical aperture is "the proportion of the impinging light that can be accepted and used." Figure 22.10 shows a fibre accepting light from a source. It should

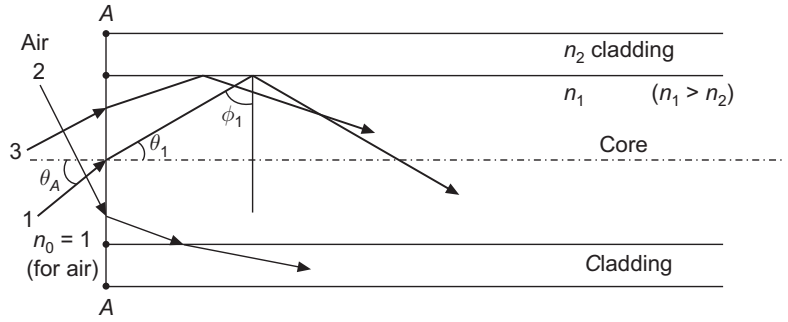


Figure 22.10 Numerical aperture and acceptance angle (longitudinal section of the fibre).

be noted that some rays from the source will not be guided because they enter the fibre at too great an incident angle. (The incident angle is the angle made by the entering ray with the axis of the core.) Their incident angle is refracted in the core and hit the core/cladding interface at an angle less than the fibre's critical angle. The interface between the outside air and the fibre glass core and cladding does not have critical angle (Section AA) because in this case $n_1 < n_2$. There are two types of rays which will propagate in the fibre i.e. meridian and skew. Meridional rays are

confined to meridian planes which are longitudinal sections of the fibre, each containing the central axis (the section shown in Figure 22.10 being one such plane). Skew rays are not restricted to one such plane and follow a helical path along the guide. Skew rays probabley make up the larger proportion of the total number of rays, but meridian rays are responsible for longer distance transmission in the fibre.

In Figure 22.10, the ray (1) whose angle of incidence on the air/core interface is θ_A is a meridian ray lying in the meridian plane coinciding with the plane of the paper. The ray (2) is refracted out of the core into the cladding. The ray (3) is a skew ray whose point of incidence is not on the line AA' (as seems to be in the figure) but lies in the plane through AA normal to the plane of the paper and its projection on the line AA is as shown in the figure. Its plane is NOT a meridian plane and the ray is skew relative to the central axis of the core.

For the critical angle ϕ_c of the core

$$\sin \phi_c = \frac{n_2}{n_1} \tag{22.14}$$

and

$$\cos^2 \phi_c = 1 - \frac{n_2^2}{n_1^2} = \frac{n_1^2 - n_2^2}{n_1^2} \tag{22.15}$$

By Snell's law $$n_0 \sin \theta_A = n_1 \sin \theta_1$$

where ϕ_A = acceptance angle and $\theta_1 = \pi/2 - \phi_1$ (22.16)

Since $n_0 = 1$ for air outside the core,

and $$\sin\theta_1 = \cos\phi_1 = \cos\phi_c$$

$$\therefore \quad \left.\begin{aligned} \sin \theta_A &= n_1 \cos \theta_c = n_1 \sqrt{\left(\frac{n_1^2 - n_2^2}{n_1^2}\right)} \\ &= \sqrt{(n_1^2 - n_2^2)} = \sqrt{\{(n_1 - n_2)(n_1 + n_2)\}} \\ &= \text{Numerical Aperture (NA)} \end{aligned}\right\} \tag{22.17}$$

Also $$\text{NA} \simeq n_1 \sqrt{2\Delta} \tag{22.17a}$$

The larger the value of NA, the greater is the fibre's ability to capture the optical power from a source and transmit it. Multimode fibres will have significantly larger NA than the single-mode fibres because of their larger cores. But they will also have modal dispersion which is absent in single-mode fibres. For the step-index fibres, NA is constant across the face of the core whereas for the graded-index core the acceptance angle varies with the distance from the core axis.

Note: The acceptance angle θ_A is the maximum angle over which the light rays entering the fibre core will be trapped in the core. It has been seen that the maximum angle occurs when ϕ_c is the critical angle satisfying the condition for total internal reflection.

22.5.2.3 Modes

Since the fibre is basically a circular waveguide, the propagation of light, which is an electromagnetic wave will be in the form of modes, as seen in Chapter 18. Also from the previous discussion in that chapter, three types of modes can exist in the fibre, i.e. the transverse electric

modes (with $E_z = 0$, $H_z \neq 0$, z being the direction of propagation — TE), the transverse magnetic modes (with $H_z = 0$ and $E_z \neq 0$ — TM), and the hybrid modes (with $E_z \neq 0$, $H_z \neq 0$—HE). It should also be remembered that the fields have circular symmetry when $E_z = 0$ or $H_z = 0$. When both E_z and H_z are non-zero, there is no circular symmetry. Another point to be noted is that all the modes excluding HE_{11} have finite cut-off frequencies. The HE_{11} mode does not have a cut-off frequency and hence the light of any wavelength can propagate inside an optical fibre in this mode.

This phenomenon gives a basis for defining the type of fibre, i.e. if only the HE_{11} mode propogates inside the fibre, then the fibre is called the single-mode fibre (SM). If other modes also propagate the fibre is called mutlimode fibre (MM). Givien below in Figure 22.11 is the *E*-field pattern in some of the lower modes of a step-index cylindrical guide. Some modes are linearly polarized and they are labelled as LP modes.

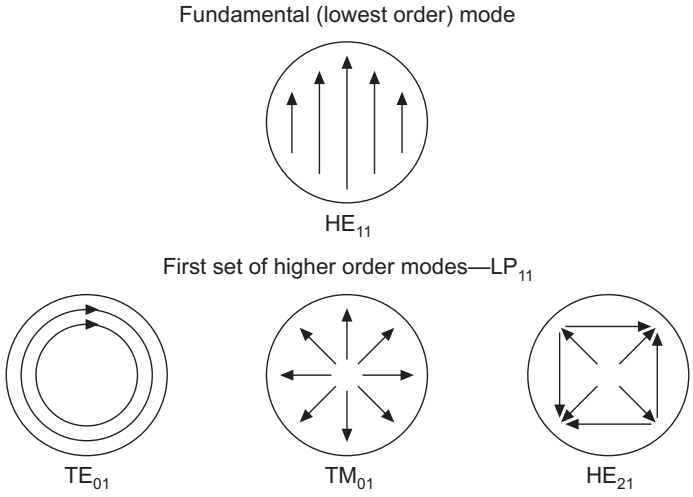


Figure 22.11 **E** field vectors in step-index waveguide lowest order modes.

The modal performance in the core of a cylindrical, step-index, optical fibre depends on a number of parameters, i.e.

(a) Radius of the core = a
(b) Free space wavelength = λ
(c) Refractive index of the core = n_1
(d) Refractive index of the cladding = n_2.

A parameter called the 'normalized frequency (=V)' is a function of all these variables and is used for a simplified analysis of fibre design. V is defined as:

$$\left.\begin{aligned} V = \frac{2\pi a}{\lambda}(n_1^2 - n_2^2)^{1/2} &= \frac{2\pi a n_1}{\lambda}\sqrt{2\Delta} \\ &= \text{normalized frequency} \\ &= \frac{\omega a}{c}(n_1^2 - n_2^2)^{1/2} \end{aligned}\right\} \tag{22.18}$$

A mode chart (or a b–V diagram) plots the relationship between these modes as a function of V and the refractive index n (or β the modal phase constant where $\beta = \omega n/c$) between the limits n_1 and n_2 (the indices of core and cladding respectively, i.e. $\beta_1 = \omega n_1/c$ to $\beta_2 = \omega n_2/c$).

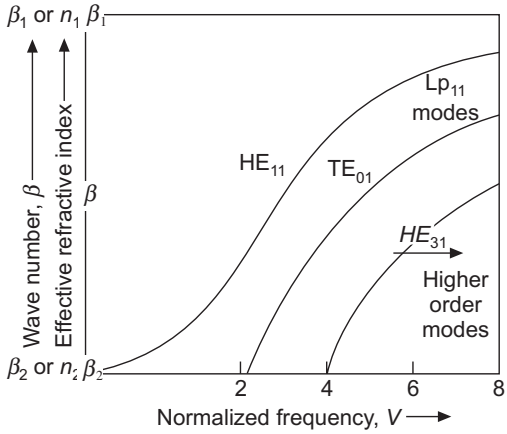


Figure 22.12 Mode chart or b-V diagram of an optical fibre (step-index).

The modal phase constant (i.e. the wave number) always lies between β_2 and β_1. As $\beta \to \beta_2$ the mode approaches the cut-off and the fields spread in the cladding. As the normalized frequency V increases, β also increases monotonically approacing asymptotically to β_1, i.e. as V increases the fields get more confined to the core. The LP_{11} mode group (i.e. TE_{01} mode) will not propagate if $V < 2.40$, which is the cut-off wavelength, and decides the region of single-mode propagation. A large difference in modal speeds would yield significant modal dispersion at the detector. The step-index modal performance can be improved by shaping the core's refractive index in the form of a parabola. This type of multimode fibre is referred to as graded index (or GRIN) fibre.

A mathematical model for a parabolic core profile of the graded index multimode fibre is given by

$$\left.\begin{aligned} n(r) &= n_1\sqrt{\left\{1-2\left(\frac{r}{a}\right)^2\Delta\right\}} && \text{for } r < a \text{ and} \\ &= n_2 && \text{for } r \geq a \end{aligned}\right\} \tag{22.19}$$

where

a = radius of the core

r = radial distance from the core-axis.

An equation for the variation of NA for this fibre, as function of the radius is

$$\text{NA} = n_1\sqrt{2\Delta}\sqrt{\left\{1-\left(\frac{r}{a}\right)^2\right\}} \tag{22.20}$$

When $$r = 0, \quad \text{NA} = n_1\sqrt{2\Delta} \tag{22.20a}$$

which is same as for the step-index type core.

The normalized frequency parameter V for a GRIN fibre is obtained using the same equation as the step-index fibre. The number of modes in a multimode GRIN fibre is reduced and approximated by

$$N = \frac{V^2}{4} \qquad \text{for } V > 10 \tag{22.21}$$

which is half the number obtained in a step-index fibre with same V.

22.5.3 Optical Fibre Performance

Though it is commonly thought that optical fibres do not have any degradation of transmission, there are three basic phenomena governing the performance of the fibres. They are:

1. Attennation losses: diminished levels of light at the optical detector due to loss of signal energy.
2. Dispersion: time of arrival differences between different wavelength components of the signal (i.e. delay distortion or dispersion)
3. Inability of conventional silica fibres to maintain polarization

22.5.3.1 Attenuation

There are three sources of losses causing the attenuation of signals. They are:

(a) *Material absorption:* Small amounts of light get absorbed in an optical fibre due to the chemical composition of glass. The energy gets converted to heat. The basic mechanism which causes this is the excitation of molecular modes of vibration called resonance. Some of the other sources are the metal impurities, which have now been controlled much better, and the water vapour in the form of OH ions. Modern manufacturing techniques have been able to control this component as well.

(b) *Scattering losses:* Light propagating in a fibre can be converted into unbound and back-scattered light due to molecular level irregularities within or on the surface of the glass. This type of attenuation is called Rayleigh scattering. The energy of this scattered light is proportional to $1/\lambda^4$.

(c) *Bending losses:* Bending in a fibre can lead to increased losses.

Large scale bends having radii greater than the fibre diameter are called macrobends and are caused by pulling, sequeezing, bending of the cable during installation. Microbends are a continuous succession of very small bends due to non-uniformity at the cladding/coating interface, non-uniform lateral pressure from cabling process, microscopic variations in the location of the core-axis.

These losses can be minimized during fibre/cable manufacturing.

22.5.3.2 Dispersion (Arrival time distortion)

Dispersion is defined as "a spreading in time of arrival of a received signal beyond its original time spread."

Each of the components in a fibre system contributes to dispersing or spreading out the signal energy over time as it (the signal) travels. A system which has been designed according to time of arrival distortion caused by dispersion is called bandwidth or rise-time limited. The fibre itself is generally the greatest contributor to the total system dispersion.

In the present generation of optical systems where only intensity (power) is modulated and detected, phase distortions cause non-constant group velocity (= v_{gr}) resulting in arrival time distortions. Pulse spreading with time and distance causes interference with each other and wavelengths also get modulated.

In multimode fibres there are many modes (hundreds of modes) present during the signal propagation in the steady state. The dominant cause of arrival time distortion in those fibres is the modal dispersion (also called inter-modal dispersion) which is the spread in arrival times of these modes. Single-mode fibres, propagating only the fundamental mode, do not have modal dispersion, i.e. material and waveguide dispersion. When combined they are called chromatic dispersion or intramodal dispersion. These effects are also present in multimode fibres but are minor relative to the modal dispesion. The effects of chromatic dispersion on the signals are proportional to the source line-width. Modal dispersion effects are not dependent on the source line-width. The intermodal dispersion is defined as the pulse broadening per unit length of the fibre and is expressed as

$$D_{\text{inter}} = \frac{n_1(n_1 - n_2)}{n_2 c} \tag{22.22a}$$

D_{inter} has unit ps/km.

The intramodal dispersion is defined as the pulse broadening per unit length of the fibre per unit spectral width of the signal and is expressed as

$$D_{\text{intra}} = \frac{\lambda V}{2\pi c}\frac{\partial^2 \beta}{\partial V^2} \tag{22.22b}$$

where D_{intra} has unit ps/km/nm.

The pulse broadening,

$$\tau = D_{\text{intra}}\, \sigma_\lambda L \tag{22.23}$$

where

σ_λ = the spectral width of the signal

L = length of the fibre.

The maximum data rate which the fibre can support is = $\frac{1}{\tau}$

$\therefore$ For higher data rates, τ and the dispersion must be as small as possible.

$\therefore$ The SM fibre has much larger bandwidth ($\because$ $D_{\text{inter}} >> D_{\text{intra}}$)

But to make a fibre single-mode, the normalized frequency of the fibre, i.e. V must be ≤ 2.4. This numerical value is a consequence of the root of the J_0. Bessel function which comes in the mathematical expressions of the cylindrical waveguide (Ref Chapter 18, Section 18.3.6.2).

(*Note:* Refering to Figure 22.12 the slope of the *b-V* curve is proportional to $\frac{\partial \beta}{\partial \omega}$ (since $V \propto \omega$).

Since $\frac{\partial \beta}{\partial \omega} = \left(\frac{\partial \omega}{\partial \beta}\right)^{-1} = \frac{1}{v_{gr}}$. Refer to Section 18.3.4, Eq. (18.100), the slope of the *b-V* curve is a measure of the group-delay time.)

22.6 LOW FREQUENCY (POWER FREQUENCY) APPLICATIONS

22.6.1 Introduction

The applications of the electromagnetic phenomena to the development of low frequency (i.e. mostly power frequency) devices started with the discovery of Faraday's law of electromagnetic induction, even before Maxwell had completed the unification of the laws of electromagnetism. The rotating energy converting devices (i.e. rotary electric motors and generators) and the subsequent allied generating equipment like trnasformers, switchgears, etc. are some of the earlier examples of applications of the electromagnetic phenomena in the power frequency range. These devices have been in use for quite some time now in different areas of applications and are quite familiar to most of us. But there are new and not so new areas of applications which are coming up with the developments in other fields such that what was not viable earlier, are becoming now possible and useful. We shall briefly outline some of these new areas here. The basic elements of the underlying theory have already been discussed in various chapters earlier (i.e. Chapters 6–15) and hence will be referred to at relevant places.

We will now list some of the important areas of application. Though major areas have been covered, it is not a fully comprehensive list and there may be areas which may not appear important at present, but can be important in the future due to discoveries in other allied areas:

1. Linear transportation—both high speed long distance as well as low speed short distance.
2. Linear pumping—linear flat as well as circular linear pumping.
3. Induction heating
4. Induction melting
5. Induction stirring
6. Controlling the flow of molten metal (i.e. steel in continuous casting of iron).
7. Transportation of metal sheets by levitation.
8. Sliding door openers.
9. Curtain openers
10. Magnetic gears
12. Electromagnetic launchers

- MHD power generation
- Superconducting motors and generators.
- Superconductivity in linear transportation.
- Electromagnetic flowmeters.
- Electromagnetic separator
- Magnetic cranes.

22.6.2 Linear Transportation

Though human mode of travel has been linear, most of the mechanical devices to aid and accelerate motion have been rotary, i.e. due to faster rotation of wheels. All mechanical devices like motor cars, trains, etc. and electrical devices like electric motors have achieved higher speeds by rotating wheels faster. The rotary motion is converted into linear motion by bringing into contact the edge of the wheel with the linear path. This process involves the use of friction

between the two surfaces because if there was no adhesion between the two surfaces (i.e. the track and wheel-edge) there would be no smooth rotation and instead the wheel would slip on the track. In the case of electrically driven wheels of the electric train, the drive motor (mostly either dc or ac induction-type), the axle is rotated by the rotating shaft of the motor. The shaft rotation is produced by the motor electromagnetically.

The electric motor (or the generator) is basically a double cylindrical structure (Figure 22.13) consisting of an inner shaft-mounted cylinder and an outer cylindrical annulus mounted co-axially with the inner cylinder such that there is an air-gap between the two which is also of cylindrical

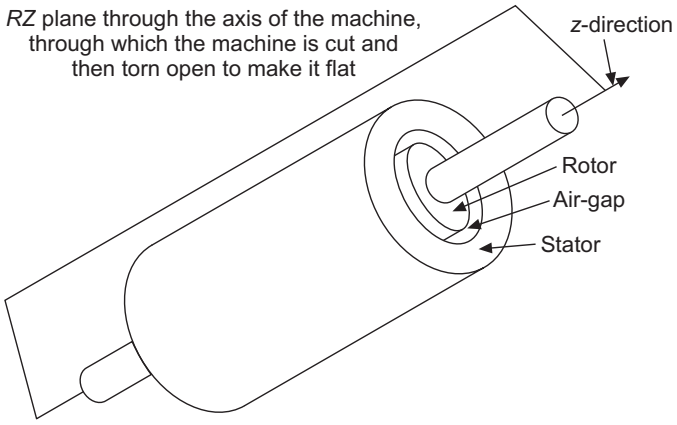

RZ plane through the axis of the machine, through which the machine is cut and then torn open to make it flat
z-direction
Rotor
Air-gap
Stator

Figure 22.13 Schematic of the basic structure of a rotating machine (not to scale). Slots and windings are not shown. Also the end-structure has been omitted. The magnetic field in the air-gap is radial, the current flow in the conductors located in the slots is axial. The effective current sheet has axially flowing currents. The magnetic field in the air gap is rotating circumferentially. The air gap magnetic field and the rotor currents interact to produce a force on the rotor such that it also rotates in the same direction as the magnetic field.

annular shape. Both are magnetic in nature (i.e. made of Si-Fe sheets stacked together) and they carry windings for carrying currents, the windings being located near the annular air-gap between the two members. The inner cylinder called the rotor is the rotating member and the outer member which is stationary is called the stator. When the device is used for converting mechanical energy to electrical energy, it is called a generator and when it converts electrical energy to mechanical energy it is called the electric motor. The energy conversion takes place by the interaction of the travelling (or rotating) magnetic fields, produced by the currents (or effective current-sheets) in the two sets of windings, across the air-gap between the two members. "When the device is acting as an electrical generater, a 3-phase current is induced in the stator winding by the rotating magnetic field in the air-gap produced by the currents in the rotor winding." The direction of the magnetic field is radial and its direction of travel is circumferential. The winding on the rotor cylinder is mechanically made to rotate so that there is a relative motion between the magnetic field in the air-gap and the stator winding, whereby an emf is induced in the winding which feeds a current into any electrical load connected to the winding (ref: Section 10.7.6 of Chapter 10, Sections 11.6 to 11.16.1, Chapter 11). When the device is operating as a motor, the rotor winding is fed with a current (direct or alternating,

depending on the type of motor) and this current in the winding interacts with the air-gap magnetic field produced by the stator currents whereby a force is generated on the rotor winding, which then makes the rotor to rotate in the same direction as the travelling magnetic field. This in brief is a qualitative description of the basic principle of operation of an electric generator or motor. The motor is connected with the shaft of the wheels which then rotate and the train thus runs on its iron tracks. Thus the rotational motion of the motor is converted into the linear motion of the train on its iron tracks.

However it is possible to convert the rotary motion of the electric motor to linear motion directly without using the intermediate device of wheels. This can be better explained by looking at Figures 22.14 and 22.15. Such devices which produce a linear force directly, are called linear motors, and as in the case of rotary motors, there are different types of linear motors, i.e. linear direct current motor, linear induction motor, linear synchronous motor and so on. The most commonly used types for linear transportation are the linear induction motor and the linear synchronous motor. Since, in these types of motors, both the stator and the rotor are flat and parallel (and not co-axially rotating), to get the advantage of direct generation of linear force and its utilization, it is essential that the moving object (which would be the 'train' in this case) is made to levitate* to eliminate the loss due to friction. It should be noted that absolutely no friction is required in such a device, since wheels are now not required for the linear motion. So now the vehicles have to be both propelled as well as levitated.

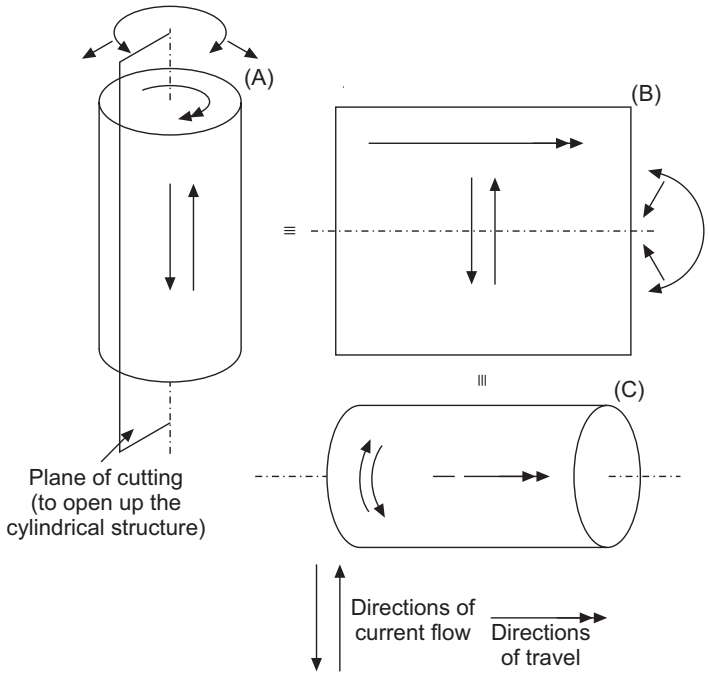


Figure 22.14 Topological manipulations to convert a cylindrical machine (A) to a flat linear machine (B), and then to a tubular linear machine (C). The direction of magnetic flux is radial in (A), normal to the plane of paper in (B), and again radial in (C).

**Note:* Levitate means that the object is lifted and floats in stable condition.

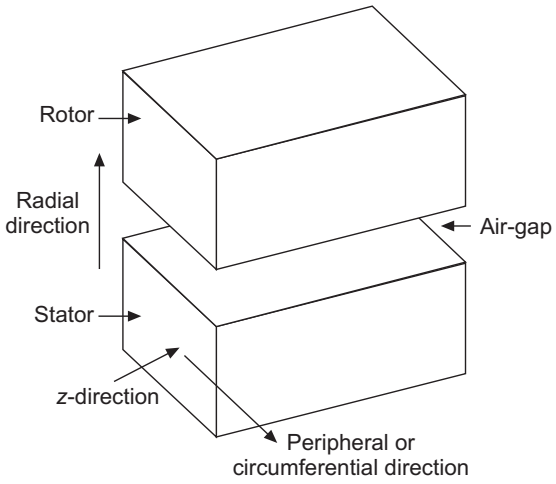


Figure 22.15 Topologically opened form of the rotating machine, converting it into flat linear machine (the dimensions are not to scale). The windings and slots have not been shown. (Basis for a single-sided flat linear machine).

The propulsion is done by the linear motor which has at least two sets of windings, i.e. a primary winding on the stator and a secondary winding on the rotor. However this arrangement is interchangable and the words stator and rotor do not have any meaning in the sense of the rotary machines, so we shall use the words related to such schemes, i.e. one set of winding on the 'track' and another set of winding on the lower side of the vehicle arranged in the slots of laminated iron sheets. Usually the three-phase primary winding is in the vehicle and the secondary winding on the track. The vehicle is fed with single-phase ac which is converted into three-phase currents and then fed to the primary windings of the linear motor. If the driving motor is a linear synhronous motor, then the track winding also has to be fed with direct current. This is done by sectionalizing the winding which makes the section of the track winding live only when the vehicle is above it. When the driving motor is a linear induction motor, the track winding is a short-circuited winding whose portions become live by the induced currents only when the moving vehicle is above a particular portion of the track.

The lifting and stabilization of the vehicle above the track can be done electromagnetically or by mechanical means, i.e. electromagnetic lifting with mechanical stabilization. The electromagnetic lifting can be of two types, i.e. (i) attractive lifting or (2) repulsive lifting. Furthermore if the train is suspended beneath the monorail type of track which is on an elevated level, then the primary windings can be laid above the vehicle and it can preferably be double-sided motor winding type. It would be clear from this brief discussion that if it is a complete electromagnetic scheme, there would be three sets of windings on the (or under the) vehicle for propulsion, supension and stabilization. In passing, it is worth mentioning that Prof. Laithwaite has suggested a type of winding which he called the 'magnetic river' where a single set of winding does all the three jobs, i.e. propulsion, lifting and stabilization. This was one of the late discoveries of Late Prof. Laithwaite, and still significant work remains to be done on this scheme before it becomes commercially viable.

At present these are linear motor driven levitated trains operating in various countries all over the world, the oldest being in the UK and then in Germany, Japan, USA, Canada, Australia and China. These are of both the types, i.e. high speed long distance as well as low speed short distance (with frequent stops). Japan has made significant progress with a high speed levitated train using

superconducting coils. The latest among the levitated trains is the one in China operating between Shanghai airport to city centre. This runs on the 'maglev principle' developed in Germany using linear synchronous motor.

22.6.3 Linear Pumps

Linear pumps are the earliest application of linear motors and were used very successfully even before the linear motors were used for transportation purposes. Though the linear pumps can be either dc or ac, the first successful pumping was done by using the principle of linear induction motor. This was done in a fast breeder reactor to transport the liquid sodium-pottassium alloy at the rate of 400 gallons per minute. The pumps were flat linear induction type (FLIP) which had two stator windings on two sides of a rectangular pipe [see Figures 22.16 and 22.17(a)] which replaced the rotor (or the secondary). Each stator had three-phase windings which produced travelling magnetic fields in the direction across the width of the pipe and travelled along the axis of the pipe (for analysis, see Problem: 11.14 of *Electromagnetism: Problems with Solutions*, 2nd Edition, 2008, in which a simplified problem neglecting the end effects and the transverse edge effects has been considered). These pumps were designed and used more because of safety considerations than their efficiency. In fact the efficiency of these pumps was only 4% but with this type of design it was possible to isolate either the hydraulic or the electrical circuit separately without breaking the other circuit.

It should be noted that as with the linear motors, the linear pump has lower efficiency compared with cylindrical devices, and the end effects are most significant compared with cylindrical devices. The reason for this is that in the cylindrical devices, there are only two ends in the axial direction. But in the flat linear devices, there are not only two axial ends, but also there are two transverse edges. Also the armature reaction is higher in flat linear devices compared to that in the cylindrical machine.

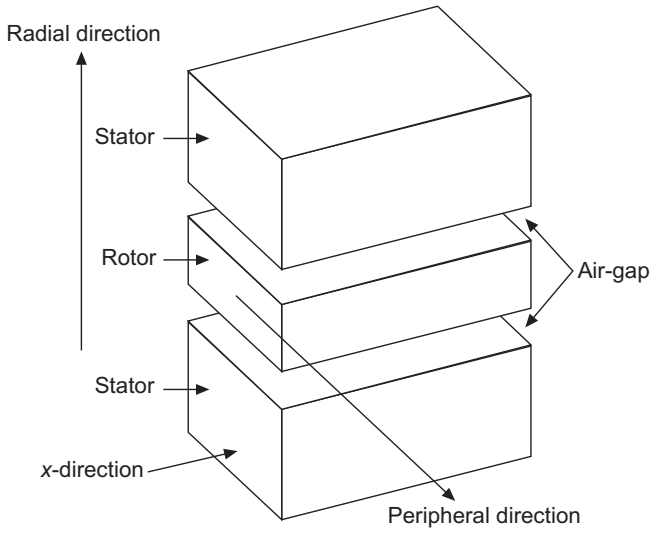
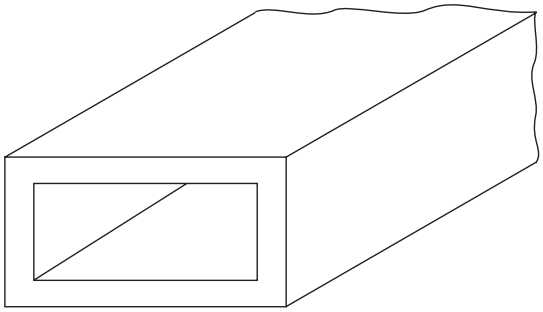
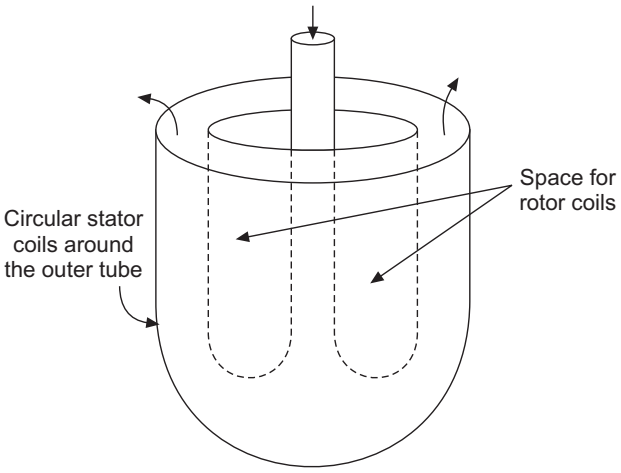


Figure 22.16 Schematic of a double-sided flat linear machine.

There was also a proposal for a tubular linear pump for larger capacity pumping of the alloy. The conceptual design of such a device can be better understood from Figure 22.14(C). A possible design of the 'rotor' of the pump is shown in Figure 22.17(b), which shows how and where the secondary coils would be located. In such a structure, both the primary and the secondary windings would be made up of circular coils which would then be arranged in a manner to produce a three-phase coil arrangement located around the outer periphery of the pump and a short-circuited secondary to be positioned in the inner annular space of the pump as shown in Figure 22.17(b). However such a pump was not constructed ultimately.



(a) Section of rectangular tube of flat linear induction pump



(b) A possible shape of a tubular linear induction pump

Figure 22.17 Rotor tubes for linear induction pumps (not to scale).

22.6.4 Induction Heating and Melting

The theory of induction heating has been adequately discussed in Chapter 15, Sections 15.2, 15.2.1, 15.3, 15.5 and 15.17 and so it will not be repeated here. Suffice it to say that induction heating of metal plates is a well-accepted practice in industry and 'surface hardening' is done by exploiting the concept of 'skin effect'.

Induction melting is merely a further extrapolation of induction heating. An arrangement for a suspended metal ball melting is shown in Figure 22.18. If this is done in vaccuo or an atmosphere

of inert gas, then highly purified liquid metal (without any oxidized surface coating, uncontaminated by any impurities in a crucible) can be obtained. The melted metal would also be suspended within the melthing-cum-levitating coil arrangement, though the shape of the molten metal changes to some extent as shown in Figure 22.18(b). For stable suspension, a short-circuited flat coil is needed to be placed at a suitable height above the crucible arrangement.

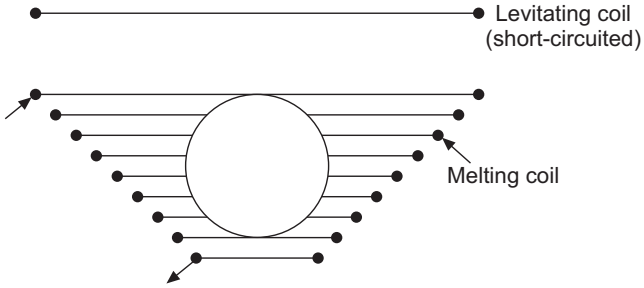


Figure 22.18(a) Induction melting of a spherical metal ball with levitation.

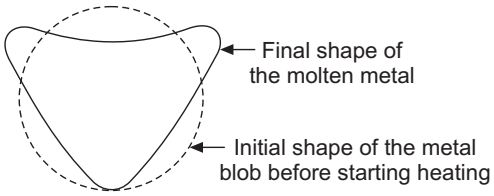


Figure 22.18(b) Change of shape of the metal blob. (The crucible coil is not shown here.)

22.6.5 Induction Stirring and Valves

Induction stirring has been widely used in steel industry and induction stirrers are commercially available. There are two types of stirrers possible, i.e. single-phase and three-phase. The three-phase stirrers which produce a travelling magnetic field are preferable as the stirring is then more uniform and evenly distributed as shown in the Figure 22.19.

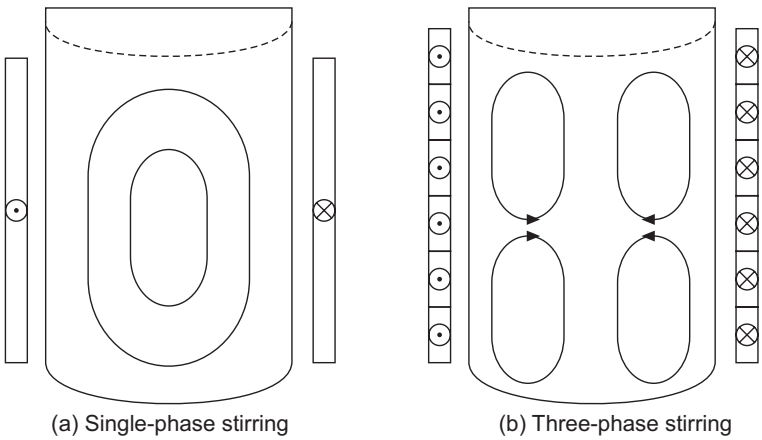


Figure 22.19 Induction stirring action in molten metal.

The use of an induction valve was suggested in the process of continuous casting of steel. When the molten alloy comes out of the bottom outlet of the tundish from where the process of continuous casting starts, it is necessary to control and stop the outflow of liquid metal at regular intervals. This is because at the tundish outlet there is ceramic pipe ot short axial length, which needs to be replaced at regular time intervals since the pipe has finite life. It was felt that during the closure periods if the metal flow could be controlled or stopped electromagnetically rather than by a mechanical device, it would be safer and cleaner. Since this pipe has circular cross-section, a tubular linear pump operating in reverse could do such a job in principle. It may have been necessary to put the secondary coils in a cylindrical enclosure, so that the molten metal would then come out of a cylindrical annulus rather than a simple cylindrical tube.

22.6.6 Transportation of Metal Sheets by Levitation

This is another direct application of linear induction motor with levitation electromagnetically. When galvanized iron sheets come out of the galvanizing bath, these sheets are transported on the rolling roller beds to the storage area. Since the sheets rub with the roller surface, there would be some scratches on these sheets. The rejection rate of these sheets is a function of the number of scratches on them. However if these sheets are levitated and propelled by linear induction motor (the sheets being the short-circuited secondary windings of the linear induction motor), the number on scratches on these sheets would be significantly reduced.

22.6.7 Magnetic Bearings and Gears

Magnetic bearings have been in use for quite some time now. Magnetic bearings come in a very wide size range, i.e. from very large ones to small ones like the suspended disc of energy meters. The bearings can be classified into two types, i.e. (a) active bearings in which balancing and positioning of the shaft is achieved by negative feedback of amplified error signals of positions of various points of the shaft and (b) passive bearings in which there is no negative feedback signal, an example being the energy meter disc.

A point to note in this matter is 'Earnshaw's theorem' according to which static equilibrium is not achievable by protential energy alone. For further information on this subject, the interested reader is recommended P.J. Geary's "Magnetic Suspension".

Magnetic gears are a relatively recent development and work is in progress on this subject in different countries. One of the attractions of magnetic gears is that the wear out of the toothed wheels in conventional gears is eliminated by these gears.

22.6.8 Electromagnetic Launchers

One of the great attractions of electromagnetic launching is that environmentally it is completely clean and there is no chemical pollution. One of the early applications of electromagnetic launchers was for short take-off of planes from the deck of aircraft carriers, though it was not a commercial success. The plane was mounted on a flat trolley with primary winding of the linear induction motor underslung from the trolley-bed and the secondary winding laid out on the take-off surface. Subsequently a lot of work has been going on for the development of electromagnetic launchers. One of the successful devices has been the rail gun which uses direct current. Electromagnetic launching of rockets, missiles, etc. requires special generators with high energy output for short durations.

22.6.9 Electromagnetic Flowmeters

A typical electromagnetic flowmeter for measuring the velocity of fluid has been discussed in Problem 6.36 of *Electromagnetism: Problems with Solutions*, 2nd Edition, 2008.

22.6.10 Electromagnetic Separator

A magnetic separator for concentrating the ore of rare earth elements was developed and successfully tested. There are different types of separators under development and in operation in a number of organizations.

22.6.11 General Comments

Other applications, mentioned in the list of Section 22.6.1, e.g. sliding door openers, curtain openers, etc. are one-off devices using small linear motors, which would use even linear dc motor or reluctance type linear motor. Another application where linear motors of small sizes have been successfully used is the 'pen drive' in X-Y plotters where the pen has to move in the *X*-direction as well as in the Y-direction. Magnetic cranes have been in use for quite some time now. Superconducting motors and generators, in spite of their early promise, have not been accepted for general use as yet. Japan has successfully experimented with high speed levitated train with superconducting windings for the linear motor and levitation. It may be commercially viable in the near future.

Rotating motors and generators have been with us for quite sometime (this application uses Faraday's law of induction) and have to be analysed for eddy current effects. There are excellent textbooks for the study and analysis of rotating machines and hence this application has not been discussed in detail here. The induction devices discussed here are mostly the devices in which 'the eddy currents' are 'usefully' used and not like rotating machines where the eddy currents are unwanted side-effects.

We have given mostly introductory comments on all these applications. Most of these devices are now being analysed in great depth by using the numerical methods like, FEM, FDM, FDTD, moment method, Green's function and so on, since the availability of modern high-speed computers. Discussion of these methods as applied to these problems would be a subject of more advanced and specalized texts beyond the scope of this book. What we have mostly discussed here, are the basics of electromagnetism from the point of view of applications.

22.7 MICROSTRIP TRANSMISSION LINES AND ANTENNAE

22.7.1 Microstrip Transmission Lines

These transmission lines belong to a group known as parallel plate transmission line and are very widely used in present-day electronics, i.e. for microwave integrated circuits, and in components like filters, couplers resonators, antennae and so on. These lines are much more flexible in comparison to co-axial cables and aslo more compact.

It consists of an infinitely large conducting plane with a metal strip placed at a distance from it (Figure 22.20) separated by a dielectric substrate. It is constructed by the photographic processes used for integrated circuits. A rigorous analytical derivation of the characteristic properties of the

line is rather tedious and cumbersome. Because of its open structure, the EM field leaks out of the dielectric substrate [Figure 22.20(b)]. An approximate formula for its characteristic impedance is

$$Z_0 = \frac{60}{\sqrt{\varepsilon_{\text{eff}}}} \ln\left(\frac{8h}{W} + \frac{W}{h}\right) \qquad \text{for} \quad \frac{W}{h} \le 1$$

and

$$\frac{1}{\sqrt{\varepsilon_{\text{eff}}}} \frac{120\pi}{\left(\frac{W}{h} + 1.393 + 0.667 \ln\left(\frac{W}{h} + 1.444\right)\right)} \qquad \text{for} \quad \frac{W}{h} \ge 1$$

where

$$\varepsilon_{\text{eff}} = \frac{\varepsilon_{\text{r}} + 1}{2} + \frac{\varepsilon_{\text{r}} - 1}{2\sqrt{\left\{1 + \frac{12h}{W}\right\}}}$$

taking account of the leakage field in air.

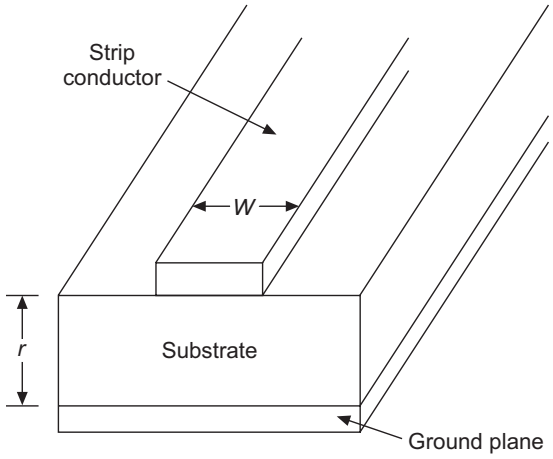


Figure 22.20(a) Microstrip transmission line.

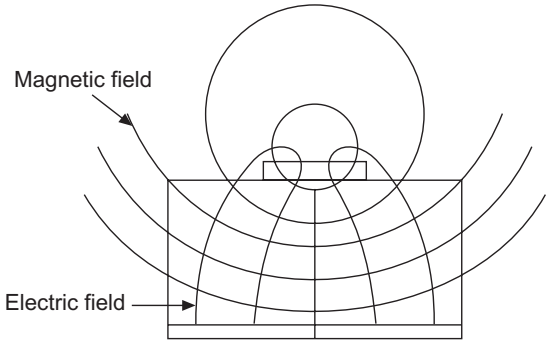


Figure 22.20(b) Pattern of electromagnetic field of a microstrip line (at any section normal to the direction of transmission).

22.7.2 Microstrip Antenna

The derivation of microstrip antenna from the transmission line came up with the aerospace applications, such as spacecraft and missiles. The basic rectangular microstrip antenna was first designed in 1972. Such a device is shown in Figure 22.21. The conducting patch shown in the figure is rectangular though it can be circular as well, or more generally it can be of any shape.

The advantages of microstrip antennae are:

(i) Low cost fabrication
(ii) Can conform to curved surface of a vehicle or product
(iii) Resistant to shock and vibration
(iv) The range of gain is considerable
(v) Antenna profile is small.

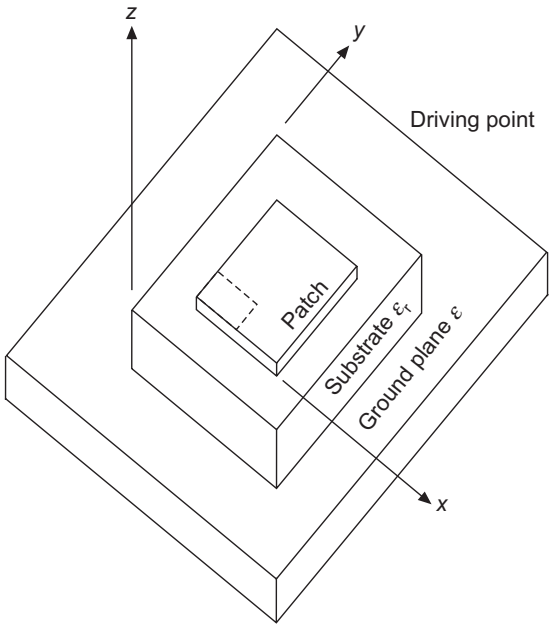


Figure 22.21 Rectangular microstrip antenna. Patch geometry is shown to be rectangular here, though it can be of any shape. The feed point is located at a point which is chosen to match the antenna with a desired impedance.

Some of the disadvantages of microstrip antennae are

(i) Narrow bandwidth
(ii) Low efficiency due to large dielectric and conductor losses
(iii) Sensitive to environmental factors like temperature and humidity.

22.7.3 Printed Antenna

Due to bandwidth requirements, in some case the antenna is designed as a 'planar antenna'. Such antennae are referred to as 'printed' antennae. Figure 22.22 shows such a 'seven-/section omni-directional antenna', which is a printed antenna.

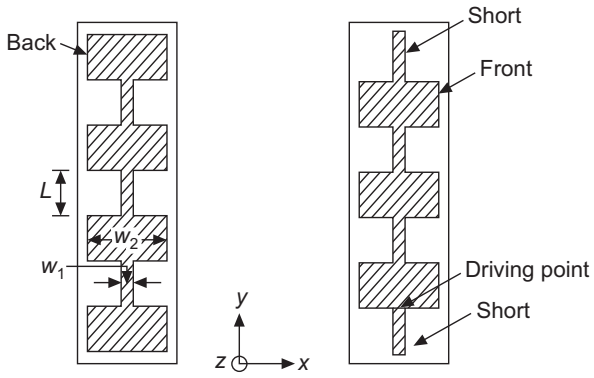


Figure 22.22 A printed antenna: seven-section omnidirectional microstrip antenna.

Appendix 1

The Principle of Duality

A.1.1 INTRODUCTION

This principle can be best explained by considering a practical example, i.e. the relation between the theory of the electric field of a linear charge and the magnetic field of a line current.

A.1.1.1 Line Charge

Figure A.1.1(a) represents the electric field of an infinitely long straight thin wire charged with $(+Q)$ coulombs per unit length.

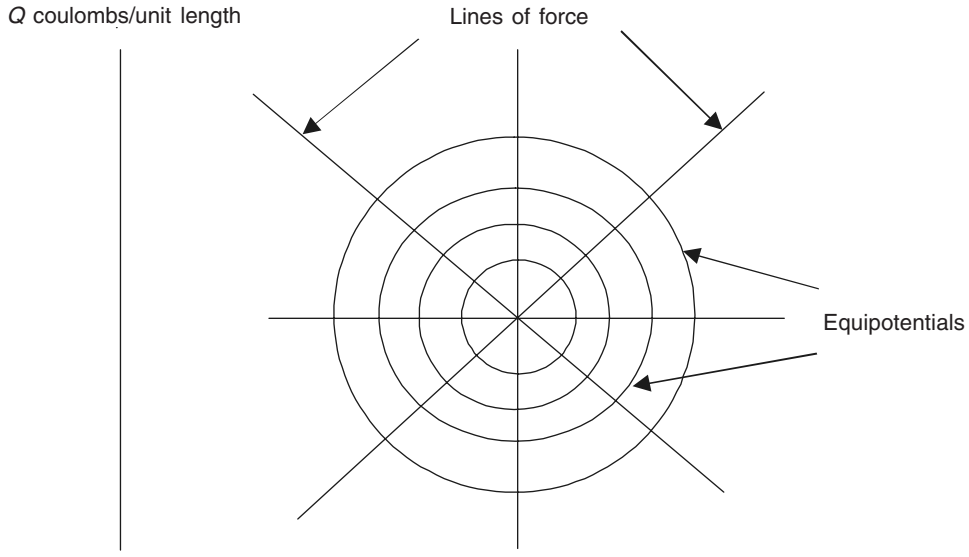


Figure A.1.1(a) Electric field of a straight, infinitely long line charge.

The electric field **E** is given by

$$E = \frac{Q}{2\pi\varepsilon r} = -\frac{\partial V}{\partial r} \tag{A.1.1}$$

$\therefore$ The potential V, by integration, is

$$V = -\left(\frac{Q}{2\pi\varepsilon}\right)\ln r \tag{A.1.2}$$

Let U be the flux function defining the lines of force. By using the conjugate relationship between U and V,

$$\left(\frac{\partial U}{\partial y}\right) = \left(\frac{\partial V}{\partial x}\right) = \left(-\frac{Q}{2\pi\varepsilon}\right)\left(\frac{\partial}{\partial x}\right)\left(\ln\sqrt{x^2 + y^2}\right)$$

$$= \left(-\frac{Q}{4\pi\varepsilon}\right)\left(\frac{2x}{x^2 + y^2}\right)$$

$$= -\frac{Qx}{2\pi\varepsilon(x^2 + y^2)}$$

$\therefore$

$$U = \left(-\frac{Qx}{2\pi\varepsilon}\right)\int\frac{dy}{x^2 + y^2}$$

$$= \left(-\frac{Q}{2\pi\varepsilon}\right)\tan^{-1}\left(\frac{y}{x}\right)$$

or

$$= \left(\frac{Q}{2\pi\varepsilon}\right)(\theta + n\pi) \tag{A.1.3}$$

It is worth noting that whilst the potential V is single-valued, the flux function is multi-valued conaining an arbitrary constant ($n\pi$). This is because even though the flux between any two lines is a definite quantity, the absolute value of the flux is indefinite since the datum from which the lines of force are reckoned, has been chosen arbitrarily. From Eq. (A.1.2), the equipotentials are given by, r = constant, i.e. concentric circles, with radii in geometric progression for equal potential drop (ΔV), i.e.

$$\Delta V = \left(\frac{Q}{2\pi\varepsilon}\right)\ln\left(\frac{r_n}{r_{n-1}}\right) \tag{A.1.4}$$

From Eq. (A.1.3), the lines of force are

θ = constant, which are equally spaced radial lines.

A.1.1.2 Line Current

Figure A.1.1(b) represents the magnetic field of an infinitely long, thin, straight wire carrying a current I amps into the plane of the paper, so that the magnetic field intensity (strength) **H** is

$$H = \frac{I}{2\pi r} = -\left(\frac{1}{r}\right)\left(\frac{\partial\Omega}{\partial\theta}\right) \tag{A.1.5}$$

where Ω is the scalar magnetic potential due to the current, and the value of **H** is given at a distance r from the axis of the wire, perpendicular to the radius and is related in its direction to that of the current by the right-hand screw rule.

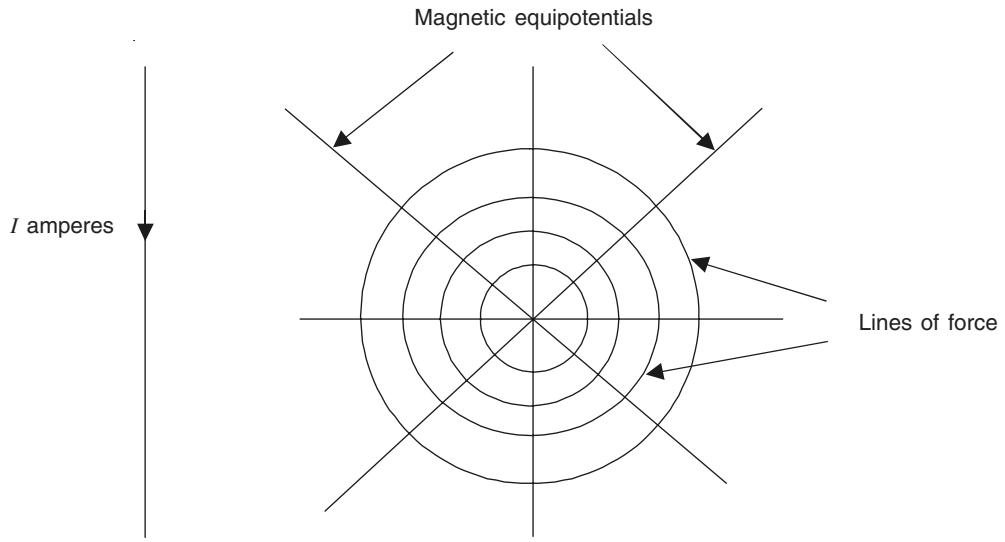


Figure A.1.1(b) Magnetic field of a straight line current.

$$\therefore \qquad \Omega = -\left(\frac{I}{2\pi}\right)(\theta + C) \tag{A.1.6}$$

The potential is thus multi-valued, and increases by an amount I which is due to the work done in moving the unit pole round the current-carrying wire. If C is chosen as $(n\pi)$, then

$$(\theta + C) = \theta + n\pi = \tan^{-1}\left(\frac{y}{x}\right) \tag{A.1.7}$$

$$\therefore \qquad \Omega = -\left(\frac{I}{2\pi}\right)\tan^{-1}\left(\frac{y}{x}\right) \tag{A.1.8}$$

The conjugate flux function Ψ is

$$\frac{\partial \Psi}{\partial y} = \frac{\partial \Omega}{\partial x} = \left(\frac{I}{2\pi}\right)\left(\frac{y}{x^2 + y^2}\right)$$

or

$$\Psi = \left(\frac{I}{2\pi}\right)\left(\frac{1}{2}\right)\ln\,(x^2 + y^2) + C_1$$

$$= \left(\frac{I}{2\pi}\right)\ln\, r + C_1 \tag{A.1.9}$$

where C_1 is arbitrary, and is fixed by the choice of the datum line of force.

The magnetic equipotentials are given by

θ = constant, which are equi-spaced radial lines.

Also, the lines of force are given by, r = constant, which are concentric circles with radii in geometrical progression.

Thus, comparing these two problems, it is seen that the potential function for the

electrostatic field of a charged wire is of the same form as the flux function of the magnetic field due to a current in the wire, and vice-versa.

The equipotentials in one problem thereupon become the lines of force in the other and vice-versa. This conjugate property of the two fields is an example of the 'Principle of Duality', according to which the electrostatic lines of force correspond with the magnetic potentials, and the electrostatic equipotentials with the magnetic lines of force for a comparable arrangement of the linear charges and currents.

By virtue of this principle, a known solution of one problem may be reciprocated into that of another problem, and conversely, simply by inverting the roles of two sets of lines and their representative functions.

Appendix 2

The Inverse Points of a Circle

A.2.1 INVERSE POINTS

We are given a circle with centre O, and one of its diameters COD (Figure A.2.1). If A and B are two points on this diameter, such that one point (say, A) is inside the circle, and the other point B is outside the circle, and they satisfy the relationship

$$OA \cdot OB = OC^2 = OP^2 \tag{A.2.1}$$

where $OC = OP$ = radius of the circle, then A and B are said to be the 'inverse points' with respect to the circle.

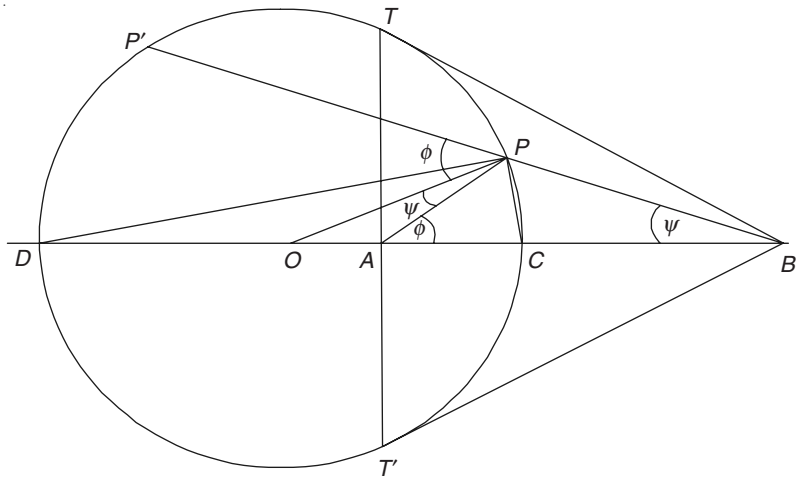


Figure A.2.1 Inverse points of a circle and the relevant angles.

The inverse points are such that, if two tangents to the circle are drawn from the external point B, i.e. BT and BT', then the chord TT' intersects the diameter DC (DCB) at A, the inverse point of B.

Geometrical Proof of (A.2.1)

Let P be a point whose locus is such that, (AP/BP) = constant.

Then, by the theorem on proportional triangles (ΔAPC, ΔBPC; and ΔADP, ΔBDP),

$$\frac{AP}{BP} = \frac{AC}{BC} = \frac{AD}{BD} \tag{A.2.2}$$

$\therefore$ For the ΔAPB, CP and DP are respectively the internal and the external bisectors of the $\angle P$, i.e. $\angle$s APB and APP' respectively.

$\therefore$ The $\angle CPD$ is a right angle.

$\therefore$ The locus of P is a circle with diameter CD.

Note: If the $\angle$s of the ΔAPB are denoted as $\angle PAB = \phi$, and $\angle PBA = \psi$, then $\angle APP' = \phi + \psi$, and $\angle DPP' = \angle DPA = \left(\frac{1}{2}\right)(\phi + \psi)$.

For the ΔAPD,

$$\angle PDA + \angle DPA + \angle PAD = \angle PDA + \left(\frac{1}{2}\right)(\phi + \psi) + (\pi - \phi) = \pi$$

$$\therefore \quad \angle PDA = \left(\frac{1}{2}\right)(\phi - \psi) \quad \text{and} \qquad \therefore \quad \angle POA = (\phi - \psi)$$

$$\therefore \quad \angle OPA = \pi - (\phi - \psi) - (\pi - \phi) = \psi \tag{A.2.3a}$$

and

$$\angle OPP' = (\phi - \psi) + \psi = \phi \tag{A.2.3b}$$

For the ΔOPA,

$$\frac{OA}{OP} = \frac{\sin OPA}{\sin OAP} = \frac{\sin \psi}{\sin (\pi - \phi)} = \frac{\sin \psi}{\sin \phi} \tag{A.2.4a}$$

And for the ΔOPB,

$$\frac{OB}{OP} = \frac{\sin OPB}{\sin OBP} = \frac{\sin [\psi + \{\pi - (\phi + \psi)\}]}{\sin \psi} = \frac{\sin \phi}{\sin \psi} \tag{A.2.4b}$$

$$\therefore \quad \frac{OA}{OP} = \frac{OP}{OB}$$

$$\therefore \quad OA \cdot OB = OP^2 = OC^2$$

Appendix 3

Green's Reciprocation Theorem in Electrostatics

A.3.1 STATEMENT OF THE THEOREM

If the charges $Q_1, Q_2, ..., Q_n$ on the conductors of a system give rise to the potentials $V_1, V_2, ..., V_n$ respectively, and if the charges $Q_1', Q_2', ..., Q_n'$ give rise to the potentials $V_1', V_2', ..., V_n'$ respectively, then:

$$\sum_{i=1}^{n} Q_i V_i' = \sum_{i=1}^{n} Q_i' V_i \tag{A.3.1}$$

Proof: We first consider a system of point charges. We then multiply one point charge by the potential of a second charge. We now write all such quantities for the n point charges in the form of a matrix of n^2 terms, as shown below. Using Eq. (1.34) of Section 1.7.1, we write down the sum of the columns in the bottom row, and the sum of the horizontal rows in the extreme right-hand column. So we have

$$\begin{array}{ccccccccccc}
0 & + & \dfrac{q_2' q_1}{4\pi\varepsilon r_{21}} & + & \dfrac{q_3' q_1}{4\pi\varepsilon r_{31}} & + & \cdots & + & \dfrac{q_n' q_1}{4\pi\varepsilon r_{n1}} & = & q_1 V_1' \\
\dfrac{q_1' q_2}{4\pi\varepsilon r_{12}} & + & 0 & + & \dfrac{q_3' q_2}{4\pi\varepsilon r_{32}} & + & \cdots & + & \dfrac{q_n' q_2}{4\pi\varepsilon r_{n2}} & = & q_2 V_2' \\
\dfrac{q_1' q_3}{4\pi\varepsilon r_{13}} & + & \dfrac{q_2' q_3}{4\pi\varepsilon r_{23}} & + & 0 & + & \cdots & + & \dfrac{q_n' q_3}{4\pi\varepsilon r_{n3}} & = & q_3 V_3' \\
\cdots & \cdots & \cdots & \cdots & \cdots & + & \cdots & + & \cdots & \cdots & \cdots \\
\dfrac{q_1' q_n}{4\pi\varepsilon r_{1n}} & + & \dfrac{q_2' q_n}{4\pi\varepsilon r_{2n}} & + & \dfrac{q_3' q_n}{4\pi\varepsilon r_{3n}} & + & \cdots & + & 0 & = & q_n V_n' \\
\hline
q_1' V_1 & + & q_2' V_2 & + & q_3' V_3 & + & \cdots & + & q_n' V_n & &
\end{array}$$

(obtained by adding the left-hand side columns)

As the order of the summation does not affect the final result, we can get the sum either by adding the terms of the last row or the last column. Equating these, we get

$$\sum_{s=1}^{n} q_s' V_s = \sum_{s=1}^{n} q_s V_s' \tag{A.3.2}$$

For these point charges, V_s is the potential at q_s due to all unprimed charges except q_s itself. All charges located on the same conductor are multiplied by the same potential and hence can be collected together. Therefore,

$$\Sigma q_i' V = V(\Sigma q_s') = Q'V \tag{A.3.3}$$

thus giving Eq. (A.3.1).

An important application of this theorem can be seen by putting

$$Q_1', Q_3', ..., Q_n' = 0$$
$$Q_2, Q_3, ..., Q_n = 0$$

and $Q_1 = Q_2'$, and make the substitution in Eq. (A.3.1).

$$\therefore \quad Q_1 V_2' = Q_2' V_2 \tag{A.3.4}$$

$$\therefore \quad V_2' = V_2 \tag{A.3.5}$$

$\therefore$ The potential to which an uncharged conductor A is raised by putting a charge Q on B is same as that to which B, when uncharged, is raised by putting a charge Q on A.

This theorem is valid, also in presence of dielectric boundaries.

A.3.2 GREEN'S RECIPROCATION THEOREM FOR DIELECTRICS

We start with Gauss' theorem for the vector **A**, i.e.

$$\sum_{j=1}^{m} \oiint_{S_j} \mathbf{A} \cdot \mathbf{n}_j \, dS_j = \iiint_v (\nabla \cdot \mathbf{A}) \, dv \tag{A.3.6}$$

where the vector **A** is a continuous function of position in space. We take the integral over a closed surface S, and the other $(m - 1)$ closed surfaces inside it, such that the integral of div **A** is defined throughout the volume v between these surfaces. Let:

$$\mathbf{A} = \psi(\varepsilon \nabla \phi) \tag{A.3.7}$$

where ψ and ϕ are scalar quantities which are finite and continuous in the region of integration and can be differentiated twice, and ε is a scalar which can be differentiated once and may be discontinuous at certain boundaries in the region. These surfaces can be excluded by drawing closely fitting surfaces on both the sides of these discontinuities. Then, if there are q such surfaces enclosing q discontinuities, adding all such contributions due to such terms Eq. (A.3.6), we have

$$\sum_{j=1}^{m} \oiint_{S_j} \varepsilon \psi \left(\frac{\partial \phi}{\partial n_j} \right) dS_j + \sum_{p=1}^{q} \oiint_{S_p} \varepsilon_p' \psi_p' \left(\frac{\partial \phi_p'}{\partial n_p'} \right) + \varepsilon_p'' \psi'' \left(\frac{\partial \phi_p''}{\partial n_p''} \right) dS_p$$
$$= \iiint_v \{\varepsilon(\nabla \phi) \cdot (\nabla \psi)\} dv + \iiint_v [\psi\{\nabla \cdot (\varepsilon \nabla \phi)\}] dv \tag{A.3.8}$$

where the single and the double dashes indicate the quantities on the two sides of the discontinuity interfaces. We can write a similar expression by interchanging ϕ and ψ, and subtracting it from (A.3.8), and thus obtain

$$\sum_{p=1}^{q} \oiint_{S_p} \left[\varepsilon'_p \left\{ \psi'_p \left(\frac{\partial \phi'_p}{\partial n'_p} \right) - \phi'_p \left(\frac{\partial \psi'_p}{\partial n'_p} \right) \right\} + \varepsilon''_p \left\{ \psi''_p \left(\frac{\partial \phi''_p}{\partial n''_p} \right) - \phi''_p \left(\frac{\partial \psi''_p}{\partial n''_p} \right) \right\} \right] dS_p$$

$$+ \sum_{j=1}^{m} \oiint_{S_j} \left[\varepsilon \left\{ \psi \left(\frac{\partial \phi}{\partial n_j} \right) - \phi \left(\frac{\partial \psi}{\partial n_j} \right) \right\} \right] dS_j$$

$$= \iiint_{v} [\psi \nabla \cdot (\varepsilon \nabla \phi) - \phi \nabla \cdot (\varepsilon \nabla \psi)] \, dv \qquad \text{(A.3.9)}$$

where n's denote the unit normals at the specific places.

In Eq. (A.3.9), let $\psi = V$ be the potential of one distribution of the charges, and $\phi = V'$ be that of another, and ε be the permittivity.

If the interface discontinuities have no charges (i.e. uncharged), then:

$$\varepsilon'_p \left(\frac{\partial \phi'}{\partial n'_p} \right) = \varepsilon''_p \left(\frac{\partial \phi''}{\partial n'_p} \right) = - \varepsilon''_p \left(\frac{\partial \phi''}{\partial n''_p} \right) \qquad \text{(A.3.10)}$$

and $\psi'_p = \psi''_p$ and $\phi'_p = \phi''_p$, and a similar set of relations for ψ's.

$\therefore$ The integrals over the surfaces of discontinuity are all zero.

If there are no charges throughout the volume,

$$\nabla \cdot (\varepsilon \nabla \psi) = 0 \qquad \text{and} \qquad \nabla \cdot (\varepsilon \nabla \phi) = 0 \qquad \text{(A.3.11)}$$

$\therefore$ The volume integral of Eq. (A.3.9) on its right-hand side = 0.

$\therefore$ From this equation, we have

$$\sum_{j=1}^{m} \oiint_{S_j} \left[V_j \varepsilon \left(\frac{\partial V'}{\partial n_j} \right) - V'_j \varepsilon \left(\frac{\partial V}{\partial n_j} \right) \right] dS_j = 0$$

or

$$\sum_{j=1}^{m} \left[V_j \oiint_{S_j} \sigma' dS_j - Vj' \oiint_{S_j} \sigma dS_j \right] = 0$$

$$\therefore \quad \sum_{j=1}^{m} Q_j V'_j = \sum_{j=1}^{m} Q'_j V_j$$

Thus the Green's Reciprocation theorem holds when the dielectrics are present.

The Reciprocation theorem can also be extended to its magnetic analogue, i.e.:

If, when a series of electrodes are kept at the potentials V_1, V_2, ..., the currents leaving them are i_1, i_2, ..., and if the potentials are changed to V'_1, V'_2, ..., the currents being i'_1, i'_2, ..., then

$$\Sigma \, iV' = \Sigma \, i'V$$

Appendix 4

Vectors in Different Coordinate Systems

A.4.1 UNIT VECTORS

A vector, whose magnitude is unity, is called a 'unit vector'. Such vectors are dimensionless quantities, used to specify directions only. The symbol **i** with various subscripts is used to denote unit vectors. The unit vector can be oriented in two different ways: (1) by specifying any two points in space and designating the vector in the direction specified by a straight line connecting one point to the other; or, (2) with reference to a coordinate system.

A.4.2 COORDINATE SYSTEMS

Many coordinate systems are possible for the definition of the location of a point in space. If at every point P in space, a coordinate system is chosen so that the coordinate curves at that point are mutually perpendicular, then the coordinates are called *orthogonal.*

In fact, there are, in use, 40 orthogonal coordinate systems, out of which only in 11 systems, simple separation of Helmholtz and Laplace's equations is possible, another 10 systems which permit R-separation of only Laplace's equation, and the remaining 18 systems permit the separation of the Laplace's equation in two dimensions only. (*Note:* The term separation is being used here to mean that the 'method of separation of variables' can be used in these coordinate systems to solve the Laplacian field problems). For further information on this topic, the interested reader is recommended to refer to *Field Theory for Engineers* and *Field Theory Handbook* by Moon & Spencer, and *Methods of Theoretical Physics* by Morse & Feshbach.

Consider the coordinate curves as shown in Figure A.4.1, where u_1, u_2, u_3 are the orthogonal coordinates. Let $\mathbf{i}_1$, $\mathbf{i}_2$, $\mathbf{i}_3$ be the unit vector tangents to the respective coordinate curves at the point P. The unit vectors chosen are positive in the directions in which the corresponding coordinates increase.

Such a coordinate system is called orthogonal curvilinear coordinate system, since the coordinate lengths are not necessarily straight lines (in general 'curves'). Let ds_1, ds_2, ds_3, be the differential components which are lengths measured as positive along the coordinate curves in the positive directions of the coordinates. The derivatives of these arcs with respect to the coordinates shall be expressed as

$$\frac{ds_1}{du_1} = h_1, \qquad \frac{ds_2}{du_2} = h_2, \qquad \frac{ds_3}{du_3} = h_3 \tag{A.4.1}$$

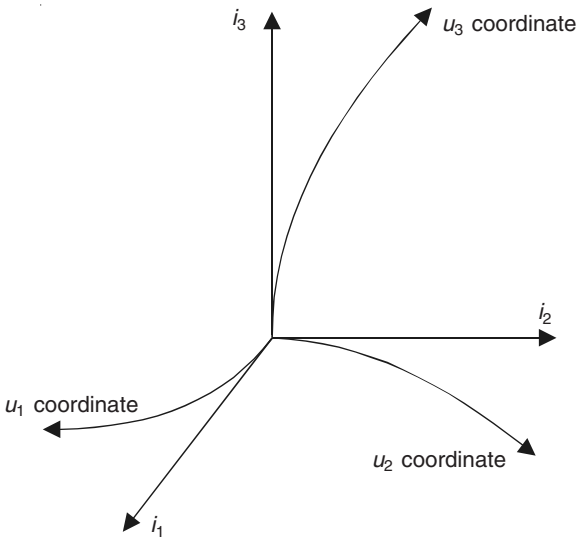


Figure A.4.1 Orthogonal curvilinear coordinates.

We remember that, in rectangular two-dimensional coordinates, in the limit

$$ds^2 = dx^2 + dy^2$$

and in two-dimensional polar coordinates (r, ϕ), the corresponding length element is

$$ds^2 = dr^2 + r^2 d\phi^2$$

So if our curvilinear coordinate system is two-dimensional, then the length element is

$$ds^2 = h_1^2 du_1^2 + h_2^2 du_2^2$$

And in a three-dimensional system, the length element will be

$$ds^2 = h_1^2 du_1^2 + h_2^2 du_2^2 + h_3^2 du_3^2 \tag{A.4.2}$$

The quantities h_1, h_2, h_3 are called the 'scale factors', for a change du_1 in the coordinate variable u_1 (say) produces a corresponding displacement $(h_1 du_1)$ along the coordinate line. In general, h can vary from point to point in space. At times these scale factors are also expressed as

$$h_1 = \sqrt{g_{11}}, \qquad h_2 = \sqrt{g_{22}}, \qquad h_3 = \sqrt{g_{33}} \tag{A.4.3}$$

where g_{11}, g_{22}, and g_{33} are called the 'metric coefficients'. It is important to remember that du_1, du_2, du_3 are not in general distances. The distances are $(h_1 du_1)$, $(h_2 du_2)$, $(h_3 du_3)$ or $\left[\left(\sqrt{g_{11}}\right)du_1\right]$, $\left[\left(\sqrt{g_{22}}\right)du_2\right]$, $\left[\left(\sqrt{g_{33}}\right)du_3\right]$.

A.4.3 DEVELOPMENT OF METRIC COEFFICIENTS

We shall now develop a formula by means of which we can calculate the metric coefficients for any coordinate system. Let us consider the system most familiar to us, i.e. a rectangular Cartesian

coordinate system (x, y, z), and a general curvilinear coordinate system (u_1, u_2, u_3). Then the functional relationship between the two systems, in general terms can be written as

$$x = x(u_1, u_2, u_3), \quad y = y(u_1, u_2, u_3), \quad z = z(u_1, u_2, u_3) \tag{A.4.4}$$

Differentiating Eqs. (A.4.4), we get

$$dx = \left(\frac{\partial x}{\partial u_1}\right) du_1 + \left(\frac{\partial x}{\partial u_2}\right) du_2 + \left(\frac{\partial x}{\partial u_3}\right) du_3 \tag{A.4.5a}$$

$$dy = \left(\frac{\partial y}{\partial u_1}\right) du_1 + \left(\frac{\partial y}{\partial u_2}\right) du_2 + \left(\frac{\partial y}{\partial u_3}\right) du_3 \tag{A.4.5b}$$

$$dz = \left(\frac{\partial z}{\partial u_1}\right) du_1 + \left(\frac{\partial z}{\partial u_2}\right) du_2 + \left(\frac{\partial z}{\partial u_3}\right) du_3 \tag{A.4.5c}$$

But in the rectangular coordinates,

$$ds^2 = dx^2 + dy^2 + dz^2 \tag{A.4.5d}$$

Hence, substituting, we have

$$\begin{aligned} ds^2 = & \left[\left(\frac{\partial x}{\partial u_1}\right)^2 + \left(\frac{\partial y}{\partial u_1}\right)^2 + \left(\frac{\partial z}{\partial u_1}\right)^2\right] du_1^2 \\ & + \left[\left(\frac{\partial x}{\partial u_2}\right)^2 + \left(\frac{\partial y}{\partial u_2}\right)^2 + \left(\frac{\partial z}{\partial u_2}\right)^2\right] du_2^2 \\ & + \left[\left(\frac{\partial x}{\partial u_3}\right)^2 + \left(\frac{\partial y}{\partial u_3}\right)^2 + \left(\frac{\partial z}{\partial u_3}\right)^2\right] du_3^2 \\ & + 2\left[\left(\frac{\partial x}{\partial u_1}\right)\left(\frac{\partial x}{\partial u_2}\right) + \left(\frac{\partial y}{\partial u_1}\right)\left(\frac{\partial y}{\partial u_2}\right) + \left(\frac{\partial z}{\partial u_1}\right)\left(\frac{\partial z}{\partial u_2}\right)\right] du_1 du_2 \\ & + 2\left[\left(\frac{\partial x}{\partial u_2}\right)\left(\frac{\partial x}{\partial u_3}\right) + \left(\frac{\partial y}{\partial u_2}\right)\left(\frac{\partial y}{\partial u_3}\right) + \left(\frac{\partial z}{\partial u_2}\right)\left(\frac{\partial z}{\partial u_3}\right)\right] du_2 du_3 \\ & + 2\left[\left(\frac{\partial x}{\partial u_3}\right)\left(\frac{\partial x}{\partial u_1}\right) + \left(\frac{\partial y}{\partial u_3}\right)\left(\frac{\partial y}{\partial u_1}\right) + \left(\frac{\partial z}{\partial u_3}\right)\left(\frac{\partial z}{\partial u_1}\right)\right] du_3 du_1 \end{aligned} \tag{A.4.6}$$

It is obvious geometrically from (Figure A.4.2), the theorem of Pythagoras, that the cross-products in the above equation must be zero, if u_1, u_2, and u_3 are mutually orthogonal.

In fact, the vanishing of the last three cross-product terms in Eq. (A.4.6) constitutes the necessary and sufficient conditions for the orthogonality.

Comparison of Eq. (A.4.6) with Eqs. (A.4.2) and (A.4.3) shows that for the orthogonal coordinate systems,

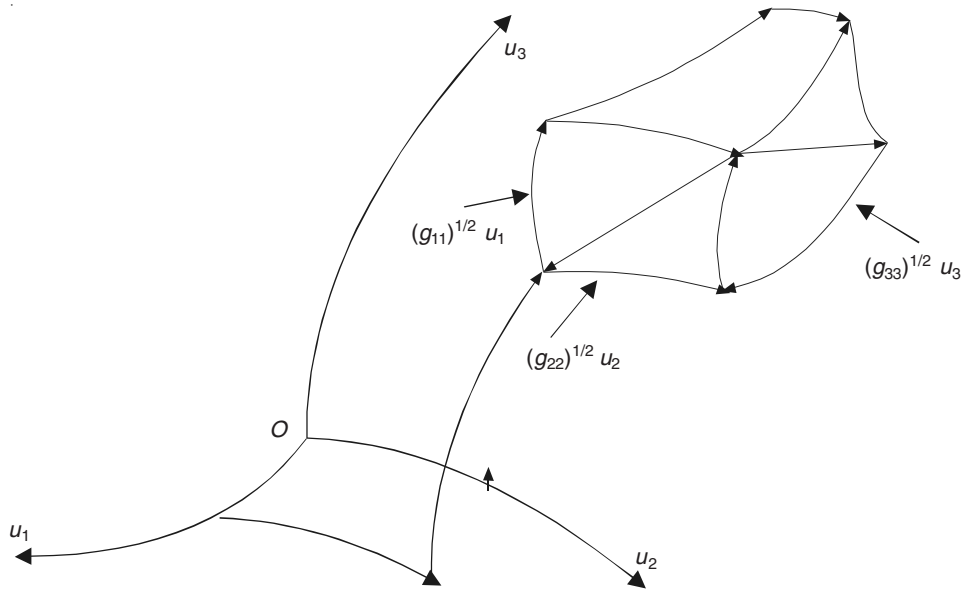


Figure A.4.2 Length elements in a general curvilinear coordinate system.

$$g_{ii} = \left(\frac{\partial x}{\partial u_i}\right)^2 + \left(\frac{\partial y}{\partial u_i}\right)^2 + \left(\frac{\partial z}{\partial u_i}\right)^2 \tag{A.4.7}$$

where $i = 1, 2, 3$.

A.4.4 TENSOR NOTATION

The expressions of the previous section can be expressed in a very simple manner if they are expressed as 'tensors' (though the knowledge of tensors is not really necessary to understand the present material).

When there are two coordinate systems (say) (u_1, u_2, u_3) and (u_1', u_2', u_3'), the relations corresponding to Eqs. (A.4.5) are

$$du_1' = \left(\frac{\partial u_1'}{\partial u_1}\right)du_1 + \left(\frac{\partial u_1'}{\partial u_2}\right)du_2 + \left(\frac{\partial u_1'}{\partial u_3}\right)du_3$$

$$du_2' = \left(\frac{\partial u_2'}{\partial u_1}\right)du_1 + \left(\frac{\partial u_2'}{\partial u_2}\right)du_2 + \left(\frac{\partial u_2'}{\partial u_3}\right)du_3$$

$$du_3' = \left(\frac{\partial u_3'}{\partial u_1}\right)du_1 + \left(\frac{\partial u_3'}{\partial u_2}\right)du_2 + \left(\frac{\partial u_3'}{\partial u_3}\right)du_3$$

or

$$du_i' = \sum_{i=1}^{3}\left(\frac{\partial u_i'}{\partial u_i}\right)du_i \tag{A.4.8a}$$

where $i = 1, 2, 3$, and $i' = 1, 2, 3$. The 'Einstein summation convention' omits the summation (Σ) when the same index is used above and below. Thus the three equations can be written very simply as

$$du_i' = \left(\frac{\partial u_i'}{\partial u_i}\right) du_i \tag{A.4.8b}$$

So corresponding to Eq. (A.4.5d),

$$ds^2 = g_{ij}\, du_i\, du_j \tag{A.4.9}$$

where the metric coefficients are

$$g_{ij} = \sum_{k=1}^{3} \left(\frac{\partial u_k'}{\partial u_i}\right)\left(\frac{\partial u_k'}{\partial u_j}\right) \tag{A.4.10}$$

The matrix of the metric coefficients is

$$g_{ij} = \begin{bmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{bmatrix} \tag{A.4.11}$$

and its determinant is called g

$$g = \begin{vmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{vmatrix} \tag{A.4.12}$$

The inverse metric coefficient g^{ij} (which is sometimes useful) is defined by the equation

$$g^{ij} g_{jk} = \delta_k^i \tag{A.4.13}$$

where δ_k^i is the Kronecker delta

$$\begin{aligned} \delta_k^i &= 1, \quad \text{if } i = k \\ &= 0, \quad \text{if i} \neq k \end{aligned} \tag{A.4.14}$$

For the 'orthogonal coordinate systems', $g_{ij} = 0$, if $i \neq j$, and hence

$$g_{ij} = \begin{bmatrix} g_{11} & 0 & 0 \\ 0 & g_{22} & 0 \\ 0 & 0 & g_{33} \end{bmatrix}$$

$$g^{ij} = \begin{bmatrix} \dfrac{1}{g_{11}} & 0 & 0 \\ 0 & \dfrac{1}{g_{22}} & 0 \\ 0 & 0 & \dfrac{1}{g_{33}} \end{bmatrix} \tag{A.4.15}$$

and

$$g = g_{11} g_{22} g_{33}$$

and

$$ds^2 = g_{ii}(du_i)^2 \tag{A.4.16}$$

where $g_{ii} = \sum_{k=1}^{3}\left(\frac{\partial u_k'}{\partial u_i}\right)^2$.

A.4.5 LENGTH, AREA, VOLUME

Infinitesimal distances expressed in the generalized coordinates are

$$ds_1 = h_1\,du_1 = \left(\sqrt{g_{11}}\right)du_1, \qquad ds_2 = h_2\,du_2 = \left(\sqrt{g_{22}}\right)du_2, \qquad ds_3 = h_3 du_3 = \left(\sqrt{g_{33}}\right)du_3$$

so that

$$ds^2 = g_{11}\,du_1^2 + g_{22}\,du_2^2 + g_{33}\,du_3^2 \tag{A.4.17}$$

The directions of ds_1, ds_2, ds_3 are given by the directions of the unit vectors $\mathbf{i}_1$, $\mathbf{i}_2$, and $\mathbf{i}_3$ respectively.

Similarly, the differential areas are

$$dA_3 = h_1h_2\,du_1\,du_2 = \left(\sqrt{g_{11}g_{22}}\right)du_1\,du_2 \tag{A.4.18a}$$

$$dA_1 = h_2h_3\,du_2\,du_3 = \left(\sqrt{g_{22}g_{33}}\right)du_2\,du_3 \tag{A.4.18b}$$

$$dA_2 = h_3h_1\,du_3\,du_1 = \left(\sqrt{g_{33}g_{11}}\right)du_3\,du_1 \tag{A.4.18c}$$

The differential volume

$$dv = h_1h_2h_3\,du_1\,du_2\,du_3 = \left(\sqrt{g_{11}g_{22}g_{33}}\right)du_1\,du_2\,du_3 \tag{A.4.19}$$

A.4.6 SPECIFIC ORTHOGONAL COORDINATE SYSTEMS AND ASSOCIATED UNIT VECTORS

The most common orthogonal coordinate systems are *Rectangular Cartesian*, *Cylindrical Polar*, *Spherical Polar*. In the present text, only the *right-handed* systems are used, i.e. if the unit vector $\mathbf{i}_1$ is rotated into $\mathbf{i}_2$, the resulting direction of advance of a right-handed screw thread establishes the direction of the third unit vector $\mathbf{i}_3$ as

$$\mathbf{i}_3 = \mathbf{i}_1 \times \mathbf{i}_2$$

A.4.6.1 Rectangular Cartesian Coordinate System

In Cartesian system of coordinates, the unit vectors may be denoted as

$\mathbf{i}_x$—unit vector in the $+x$ direction
$\mathbf{i}_y$—unit vector in the $+y$ direction
$\mathbf{i}_z$—unit vector in the $+z$ direction

The unit vectors are shown in Figure A.4.3.

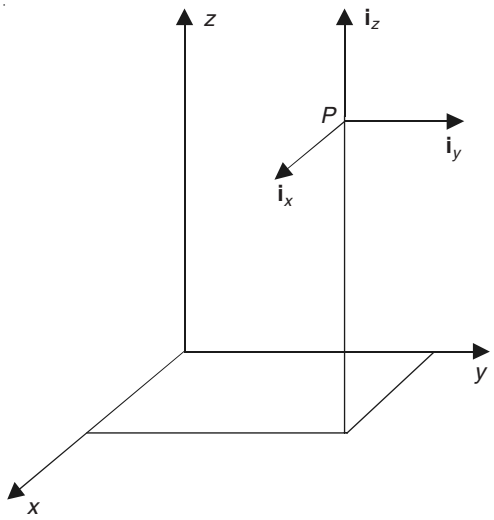


Figure A.4.3 Rectangular Cartesian coordinate system.

In the above system, the differential components of the general arc *ds* are identical with the differentials of the coordinates.

$$\therefore \quad u_1 = s_1 = x, \qquad u_2 = s_2 = y, \qquad u_3 = s_3 = z \tag{A.4.20}$$

and as a consequence:

$$h_1 = h_2 = h_3 = 1 \qquad \text{and} \qquad g_{11} = g_{22} = g_{33} = 1 \tag{A.4.21}$$

$$ds_1 = dx, \qquad ds_2 = dy, \qquad ds_3 = dz \tag{A.4.22}$$

$$dA_1 = dy\,dz, \qquad dA_2 = dz\,dx, \qquad dA_3 = dx\,dy \tag{A.4.23}$$

$$dv = dx\,dy\,dz \tag{A.4.24}$$

A.4.6.2 Cylindrical Polar Coordinate System

A radial distance (in *x*-*y* or parallel plane), an angular displacement (in the same plane as the radius above), and an axial displacement may be used to describe a cylindrical polar coordinate system which is orthogonal. Figure A.4.4 illustrates the cylindrical coordinates (r, ϕ, z), and the corresponding unit vectors in the positive directions are $\mathbf{i}_r$, $\mathbf{i}_\phi$, $\mathbf{i}_z$. Note that $\mathbf{i}_z$ is still a unit vector in the *z*-direction, and is same as the $\mathbf{i}_z$ in the Cartesian coordinate system.

From Figure A.4.4,

$$ds_1 = dr, \qquad ds_2 = d\phi, \qquad ds_3 = dz \tag{A.4.25}$$

$$\therefore \quad u_1 = r, \qquad u_2 = \phi, \qquad u_3 = z \tag{A.4.26}$$

$$du_1 = dr, \qquad du_2 = d\phi, \qquad du_3 = dz \tag{A.4.27}$$

and

$$h_1 = \frac{dr}{dr} = 1, \qquad h_2 = \frac{r\,d\phi}{d\phi} = r, \qquad h_3 = \frac{dz}{dz} = 1 \tag{A.4.28a}$$

$$\therefore \quad g_{11} = 1, \qquad g_{22} = r^2, \qquad g_{33} = 1, \qquad \sqrt{g} = r \tag{A.4.28b}$$

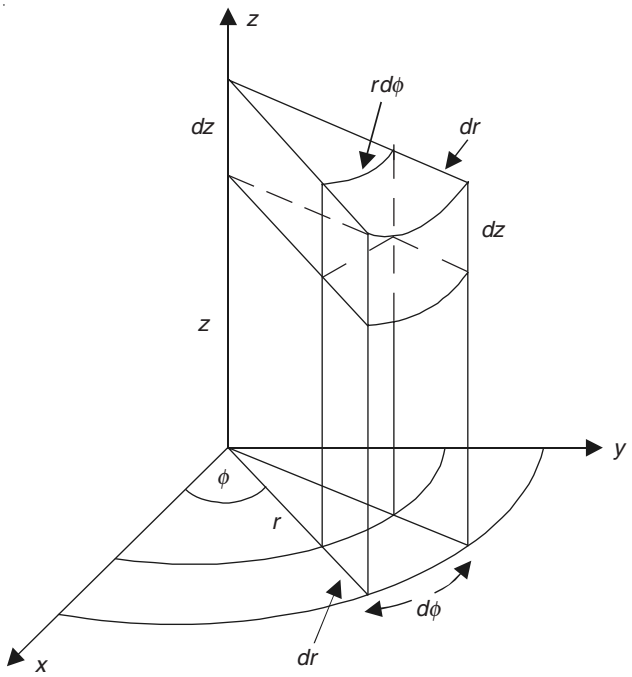


Figure A.4.4 Cylindrical polar coordinate system.

$$ds^2 = dr^2 + r^2 d\phi^2 + dz^2 \tag{A.4.29}$$

$$dA_1 = r\,d\phi\,dz, \qquad dA_2 = dz\,dr, \qquad dA_3 = r\,dr\,d\phi \tag{A.4.30}$$

$$dv = r\,dr\,d\phi\,dz \tag{A.4.31}$$

A.4.6.3 Spherical Polar Coordinate System

A radial displacement (about the origin), and two angular displacements may be used to establish a spherical coordinate system which is also orthogonal.

Figure A.4.5 illustrates the spherical polar coordinate system (ρ, ϕ, θ), and the unit vectors in the directions of the coordinate axes are $\mathbf{i}_\rho$, $\mathbf{i}_\phi$, $\mathbf{i}_\theta$. Note that $\mathbf{i}_\phi$ is same in both the cylindrical and the spherical coordinate systems.

From Figure A.4.5,

$$ds_1 = d\rho, \qquad ds_2 = \rho d\theta, \qquad ds_3 = \rho \sin\theta\, d\phi \tag{A.4.32}$$

$$\therefore \quad u_1 = \rho, \qquad u_2 = \theta, \qquad u_3 = \phi \tag{A.4.33}$$

$$du_1 = d\rho, \qquad du_2 = d\theta, \qquad du_3 = d\phi \tag{A.4.34}$$

and

$$h_1 = \frac{d\rho}{d\rho} = 1, \qquad h_2 = \left(\frac{\rho d\phi}{d\phi}\right) = \rho, \qquad h_3 = \rho \sin\theta \left(\frac{d\phi}{d\phi}\right) = \rho \sin\theta \tag{A.4.35a}$$

$$\therefore \quad g_{11} = 1, \qquad g_{22} = \rho^2, \qquad g_{33} = \rho^2 \sin^2\theta, \qquad \sqrt{g} = \rho^2 \sin\theta \tag{A.4.35b}$$

$$ds^2 = d\rho^2 + \rho^2 d\theta^2 + \rho^2 \sin^2\theta\, d\phi^2 \tag{A.4.36}$$

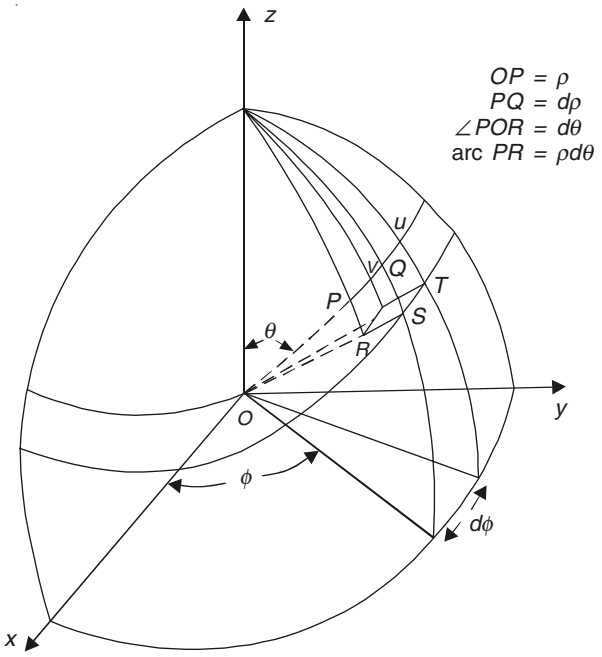


Figure A.4.5 Spherical polar coordinate system (showing the details of the volume elements).

$$dA_1 = \rho^2 \sin\theta\, d\theta\, d\phi, \qquad dA_2 = \rho \sin\theta\, d\rho\, d\phi, \qquad dA_3 = \rho\, d\rho\, d\theta \tag{A.4.37}$$

$$dv = \rho^2 \sin\theta\, d\rho\, d\theta\, d\phi \tag{A.4.38}$$

A.4.7 RADIUS VECTORS AND ELEMENTS OF LENGTHS

In Cartesian coordinates,

$$\mathbf{r} = \mathbf{i}_x x + \mathbf{i}_y y + \mathbf{i}_z z \tag{A.4.39}$$

$$d\mathbf{r} \text{ or } d\mathbf{l} = \mathbf{i}_x\, dx + \mathbf{i}_y\, dy + \mathbf{i}_z\, dz \tag{A.4.40}$$

In cylindrical coordinates,

$$\mathbf{r} = \mathbf{i}_r r + \mathbf{i}_z z \tag{A.4.41}$$

$$d\mathbf{r} \text{ or } d\mathbf{l} = \mathbf{i}_r dr + \mathbf{i}_\phi r\, d\phi + \mathbf{i}_z\, dz \tag{A.4.42}$$

In spherical coordinates,

$$\mathbf{r} = \mathbf{i}_\rho \rho \tag{A.4.43}$$

$$d\mathbf{r} \text{ or } d\mathbf{l} = \mathbf{i}_\rho\, d\rho + \mathbf{i}_\theta \rho\, d\theta + \mathbf{i}_\phi \rho \sin\theta\, d\phi \tag{A.4.44}$$

A.4.8 VECTOR OPERATOR ∇

As we have seen earlier, the vectors, in general, may be functions of both space and time. The operator 'del' (= ∇) is a vector function space operator which, when expressed in 'differential terms' can be defined through the partial derivatives with respect to space. Physically what the partial differentiation implies is that during this space operation, we are holding the instant of time fixed, while investigating their spatial behaviour.

In any orthogonal coordinate system, the vector operator ∇ is

$$\nabla = \mathbf{i}_1\left(\frac{\partial}{\partial s_1}\right) + \mathbf{i}_2\left(\frac{\partial}{\partial s_2}\right) + \mathbf{i}_3\left(\frac{\partial}{\partial s_3}\right) \tag{A.4.45}$$

or, in terms of the coordinates (u_1, u_2, u_3), since, in general, the spatial differential $ds = h\,du$,

$$\nabla = \mathbf{i}_1\left(\frac{1}{h_1}\right)\left(\frac{\partial}{\partial u_1}\right) + \mathbf{i}_2\left(\frac{1}{h_2}\right)\left(\frac{\partial}{\partial u_2}\right) + \mathbf{i}_3\left(\frac{1}{h_3}\right)\left(\frac{\partial}{\partial u_3}\right) \tag{A.4.46}$$

or, in terms of the metric coefficients,

$$\nabla = \mathbf{i}_1\left(\frac{1}{\sqrt{g_{11}}}\right)\left(\frac{\partial}{\partial u_1}\right) + \mathbf{i}_2\left(\frac{1}{\sqrt{g_{22}}}\right)\left(\frac{\partial}{\partial u_2}\right) + \mathbf{i}_3\left(\frac{1}{\sqrt{g_{33}}}\right)\left(\frac{\partial}{\partial u_3}\right) \tag{A.4.47}$$

In Cartesian coordinates,

$$\nabla = \mathbf{i}_x\left(\frac{\partial}{\partial x}\right) + \mathbf{i}_y\left(\frac{\partial}{\partial y}\right) + \mathbf{i}_z\left(\frac{\partial}{\partial z}\right) \tag{A.4.48}$$

In cylindrical coordinates,

$$\nabla = \mathbf{i}_r\left(\frac{\partial}{\partial r}\right) + \mathbf{i}_\phi\left(\frac{1}{r}\right)\left(\frac{\partial}{\partial \phi}\right) + \mathbf{i}_z\left(\frac{\partial}{\partial z}\right) \tag{A.4.49}$$

In spherical coordinates,

$$\nabla = \mathbf{i}_\rho\left(\frac{\partial}{\partial \rho}\right) + \mathbf{i}_\theta\left(\frac{1}{\rho}\right)\left(\frac{\partial}{\partial \theta}\right) + \mathbf{i}_\phi\left(\frac{1}{\rho \sin\theta}\right)\left(\frac{\partial}{\partial \phi}\right) \tag{A.4.50}$$

Note: For Cartesian coordinates,

$$h_1 = h_2 = h_3 = 1 = \sqrt{g_{11}} = \sqrt{g_{22}} = \sqrt{g_{33}} \tag{A.4.51a}$$

and

$$u_1 = x, \qquad u_2 = y, \qquad u_3 = z \tag{A.4.51b}$$

For cylindrical coordinates,

$$h_1 = 1 = \sqrt{g_{11}}, \qquad h_2 = r = \sqrt{g_{22}}, \qquad h_3 = 1 = \sqrt{g_{33}} \tag{A.4.52a}$$

and

$$u_1 = r, \qquad u_2 = \phi, \qquad u_3 = z \tag{A.4.52b}$$

For spherical coordinates,

$$h_1 = 1 = \sqrt{g_{11}}, \qquad h_2 = \rho = \sqrt{g_{22}}, \qquad h_3 = \rho \sin\theta = \sqrt{g_{33}} \tag{A.4.53a}$$

and

$$u_1 = \rho, \qquad u_2 = \theta, \qquad u_3 = \phi \tag{A.4.53b}$$

A.4.9 GRADIENT OF A SCALAR FUNCTION ($= \nabla\varphi$)

In the curvilinear coordinate system,

$$\nabla\varphi = \mathbf{i}_1\left(\frac{1}{h_1}\right)\left(\frac{\partial\varphi}{\partial u_1}\right) + \mathbf{i}_2\left(\frac{1}{h_2}\right)\left(\frac{\partial\varphi}{\partial u_2}\right) + \mathbf{i}_3\left(\frac{1}{h_3}\right)\left(\frac{\partial\varphi}{\partial u_3}\right) \tag{A.4.54}$$

In terms of the metric coefficients,

$$\nabla\varphi = \mathbf{i}_1\left(\frac{1}{\sqrt{g_{11}}}\right)\left(\frac{\partial\varphi}{\partial u_1}\right) + \mathbf{i}_2\left(\frac{1}{\sqrt{g_{22}}}\right)\left(\frac{\partial\varphi}{\partial u_2}\right) + \mathbf{i}_3\left(\frac{1}{\sqrt{g_{33}}}\right)\left(\frac{\partial\varphi}{\partial u_3}\right) \tag{A.4.55}$$

In the three orthogonal coordinate systems, this will be

Cartesian:
$$\nabla\varphi = \mathbf{i}_x\left(\frac{\partial\varphi}{\partial x}\right) + \mathbf{i}_y\left(\frac{\partial\varphi}{\partial y}\right) + \mathbf{i}_z\left(\frac{\partial\varphi}{\partial z}\right) \tag{A.4.56}$$

Cylindrical:
$$\nabla\varphi = \mathbf{i}_r\left(\frac{\partial\varphi}{\partial r}\right) + \mathbf{i}_\phi\left(\frac{1}{r}\right)\left(\frac{\partial\varphi}{\partial\phi}\right) + \mathbf{i}_z\left(\frac{\partial\varphi}{\partial z}\right) \tag{A.4.57}$$

Spherical:
$$\nabla\varphi = \mathbf{i}_\rho\left(\frac{\partial\varphi}{\partial\rho}\right) + \mathbf{i}_\theta\left(\frac{1}{\rho}\right)\left(\frac{\partial\varphi}{\partial\theta}\right) + \mathbf{i}_\phi\left(\frac{1}{\rho\sin\theta}\right)\left(\frac{\partial\varphi}{\partial\phi}\right) \tag{A.4.58}$$

(*Note:* φ is a scalar space function.)

A.4.10 DIVERGENCE OF A VECTOR ($= \nabla \cdot \mathbf{A}$)

In the curvilinear coordinate system,

$$\nabla\cdot\mathbf{A} = \left(\frac{1}{h_1h_2h_3}\right)\left(\frac{\partial}{\partial u_1}\right)(h_2h_3A_1) + \left(\frac{\partial}{\partial u_2}\right)(h_3h_1A_2) + \left(\frac{\partial}{\partial u_3}\right)(h_1h_2A_3) \tag{A.4.59}$$

or, in terms of the metric coefficients,

$$\nabla\cdot\mathbf{A} = \left(\sqrt{g}\right)\left[\left(\frac{\partial}{\partial u_1}\right)\left\{\left(\frac{g}{g_{11}}\right)^{1/2}A_1\right\} + \left(\frac{\partial}{\partial u_2}\right)\left\{\left(\frac{g}{g_{22}}\right)^{1/2}A_2\right\} + \left(\frac{\partial}{\partial u_3}\right)\left\{\left(\frac{g}{g_{33}}\right)^{1/2}A_3\right\}\right] \tag{A.4.60}$$

where

$$\mathbf{A} = \mathbf{i}_1A_1 + \mathbf{i}_2A_2 + \mathbf{i}_3A_3 \tag{A.4.61}$$

The above expressions simplify to:

Cartesian:
$$\nabla\cdot\mathbf{A} = \left(\frac{\partial A_x}{\partial x}\right) + \left(\frac{\partial A_y}{\partial y}\right) + \left(\frac{\partial A_z}{\partial z}\right) \tag{A.4.62}$$

Cylindrical:
$$\nabla\cdot\mathbf{A} = \left(\frac{1}{r}\right)\left[\left(\frac{\partial}{\partial r}\right)(rA_r) + \left(\frac{\partial}{\partial\phi}\right)(A_\phi) + \left(\frac{\partial}{\partial z}\right)(rA_z)\right]$$

$$= \left(\frac{A_r}{r}\right) + \left(\frac{\partial A_r}{\partial r}\right) + \left(\frac{1}{r}\right)\left(\frac{\partial A_\phi}{\partial\phi}\right) + \left(\frac{\partial A_z}{\partial z}\right) \tag{A.4.63}$$

Spherical:
$$\nabla\cdot\mathbf{A} = \left(\frac{1}{\rho^2\sin\theta}\right)\left[\left(\frac{\partial}{\partial\rho}\right)(\rho^2\sin\theta A_\rho) + \left(\frac{\partial}{\partial\theta}\right)(\rho\sin\theta A_\theta) + \left(\frac{\partial}{\partial\phi}\right)(\rho A_\phi)\right]$$

$$= \frac{2A_\rho}{\rho} + \left(\frac{\partial A_\rho}{\partial\rho}\right) + \left(\frac{1}{\rho}\right)\left(\frac{\partial A_\theta}{\partial\theta}\right) + \left(\frac{\cot\theta}{\rho}\right)A_\theta + \left(\frac{1}{\rho\sin\theta}\right)\left(\frac{\partial A_\phi}{\partial\phi}\right) \tag{A.4.64}$$

A.4.11 CURL OF A VECTOR (= $\nabla \times \mathbf{A}$)

In the curvilinear coordinate system, the curl of a vector can be expressed in differential form, as the following determinant,

$$\nabla \times \mathbf{A} = \left(\frac{1}{h_1 h_2 h_3}\right) \begin{vmatrix} \mathbf{i}_1 h_1 & \mathbf{i}_2 h_2 & \mathbf{i}_3 h_3 \\ \dfrac{\partial}{\partial u_1} & \dfrac{\partial}{\partial u_2} & \dfrac{\partial}{\partial u_3} \\ h_1 A_1 & h_2 A_2 & h_3 A_3 \end{vmatrix} \tag{A.4.65}$$

or expressing in terms of the metric coefficients,

$$\nabla \times \mathbf{A} = \left(\frac{1}{\sqrt{g}}\right) \begin{vmatrix} \mathbf{i}_1 \left(\sqrt{g_{11}}\right) & \mathbf{i}_2 \left(\sqrt{g_{22}}\right) & \mathbf{i}_3 \left(\sqrt{g_{33}}\right) \\ \dfrac{\partial}{\partial u_1} & \dfrac{\partial}{\partial u_2} & \dfrac{\partial}{\partial u_3} \\ \left(\sqrt{g_{11}}\right) A_1 & \left(\sqrt{g_{22}}\right) A_2 & \left(\sqrt{g_{33}}\right) A_3 \end{vmatrix} \tag{A.4.66}$$

This simplifies to the following in the three orthogonal coordinate systems:

Cartesian:
$$\nabla \times \mathbf{A} = \begin{vmatrix} \mathbf{i}_x & \mathbf{i}_y & \mathbf{i}_z \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ A_x & A_y & A_z \end{vmatrix} \tag{A.4.67}$$

Cylindrical:
$$\nabla \times \mathbf{A} = \left(\frac{1}{r}\right) \begin{vmatrix} \mathbf{i}_r & \mathbf{i}_\phi & \mathbf{i}_z \\ \dfrac{\partial}{\partial r} & \dfrac{\partial}{\partial \phi} & \dfrac{\partial}{\partial z} \\ A_r & r A_\phi & A_z \end{vmatrix} \tag{A.4.68}$$

Spherical:
$$\nabla \times \mathbf{A} = \left(\frac{1}{\rho^2 \sin\theta}\right) \begin{vmatrix} \mathbf{i}_\rho & \mathbf{i}_\theta \rho & \mathbf{i}_\phi \rho \sin\theta \\ \dfrac{\partial}{\partial \rho} & \dfrac{\partial}{\partial \theta} & \dfrac{\partial}{\partial \phi} \\ A_\rho & \rho A_\theta & \rho \sin\theta A_\phi \end{vmatrix} \tag{A.4.69}$$

A.4.12 (SCALAR) LAPLACIAN ∇^2 = DIVERGENCE OF THE GRADIENT OF A SCALAR FUNCTION φ

From Eq. (A.4.54), in the curvilinear coordinates,

$$\text{div grad } \varphi = \nabla \cdot \nabla\varphi = \nabla^2\varphi$$

$$= \left(\frac{1}{h_1h_2h_3}\right)\left[\left(\frac{\partial}{\partial u_1}\right)\left\{\left(\frac{h_2h_3}{h_1}\right)\left(\frac{\partial\varphi}{\partial u_1}\right)\right\}\right.$$

$$\left. + \left(\frac{\partial}{\partial u_2}\right)\left\{\left(\frac{h_3h_1}{h_2}\right)\left(\frac{\partial\varphi}{\partial u_2}\right)\right\} + \left(\frac{\partial}{\partial u_3}\right)\left\{\left(\frac{h_1h_2}{h_3}\right)\left(\frac{\partial\varphi}{\partial u_3}\right)\right\}\right] \tag{A.4.70}$$

Or in terms of metric coefficients,

$$\nabla^2\varphi = \left(\frac{1}{\sqrt{g}}\right)\left[\left(\frac{\partial}{\partial u_1}\right)\left\{\left(\frac{\sqrt{g}}{g_{11}}\right)\left(\frac{\partial\varphi}{\partial u_1}\right)\right\}\right.$$

$$\left. + \left(\frac{\partial}{\partial u_2}\right)\left\{\left(\frac{\sqrt{g}}{g_{22}}\right)\left(\frac{\partial\varphi}{\partial u_2}\right)\right\} + \left(\frac{\partial}{\partial u_3}\right)\left\{\left(\frac{\sqrt{g}}{g_{33}}\right)\left(\frac{\partial\varphi}{\partial u_3}\right)\right\}\right] \tag{A.4.71}$$

Either of these two expressions will reduce to the following in the three orthogonal coordinate systems:

Cartesian: $$\nabla^2\varphi = \left(\frac{\partial^2\varphi}{\partial x^2}\right) + \left(\frac{\partial^2\varphi}{\partial y^2}\right) + \left(\frac{\partial^2\varphi}{\partial z^2}\right) \tag{A.4.72}$$

Cylindrical: $$\nabla^2\varphi = \left(\frac{1}{r}\right)\left[\left(\frac{\partial}{\partial r}\right)\left(\frac{r\partial\varphi}{\partial r}\right) + \left(\frac{\partial}{\partial\phi}\right)\left\{\left(\frac{1}{r}\right)\left(\frac{\partial\varphi}{\partial\phi}\right)\right\} + \left(\frac{\partial}{\partial z}\right)\left(\frac{r\partial\varphi}{\partial z}\right)\right]$$

$$= \left(\frac{1}{r}\right)\left(\frac{\partial\varphi}{\partial r}\right) + \left(\frac{\partial^2\varphi}{\partial r^2}\right) + \left(\frac{1}{r^2}\right)\left(\frac{\partial^2\varphi}{\partial\phi^2}\right) + \left(\frac{\partial^2\varphi}{\partial z^2}\right) \tag{A.4.73}$$

Spherical: $$\nabla^2\varphi = \left(\frac{1}{\rho^2\sin\theta}\right)\left[\left(\frac{\partial}{\partial\rho}\right)\left\{(\rho^2\sin\theta)\left(\frac{\partial\varphi}{\partial\rho}\right)\right\}\right.$$

$$\left. + \left(\frac{\partial}{\partial\theta}\right)\left\{\sin\theta\left(\frac{\partial\varphi}{\partial\theta}\right)\right\} + \left(\frac{\partial}{\partial\phi}\right)\left\{\left(\frac{1}{\sin\theta}\right)\left(\frac{\partial\varphi}{\partial\phi}\right)\right\}\right]$$

$$= \left(\frac{2}{\rho}\right)\left(\frac{\partial\varphi}{\partial\rho}\right) + \left(\frac{\partial^2\varphi}{\partial\rho^2}\right) + \left(\frac{\cot\theta}{\rho^2}\right)\left(\frac{\partial^2\varphi}{\partial\theta^2}\right) + \frac{1}{\rho^2\sin^2\theta}\left(\frac{\partial^2\varphi}{\partial\phi^2}\right) \tag{A.4.74}$$

A.4.13 RELATIONS BETWEEN THE THREE ORTHOGONAL COORDINATE SYSTEMS

A.4.13.1 Cartesian (x, y, z) and Cylindrical (r, ϕ, z)

$$x = r\cos\phi, \qquad y = r\sin\phi, \qquad z = z \tag{A.4.75a}$$

$$r^2 = x^2 + y^2, \qquad \phi = \tan^{-1}\left(\frac{y}{x}\right), \qquad z = z \tag{A.4.75b}$$

A.4.13.2 Cartesian (x, y, z) and Spherical (ρ, θ, ϕ)

$$x = \rho \sin\theta \cos\phi, \qquad y = \rho \sin\theta \sin\phi, \qquad z = \rho \cos\theta \tag{A.4.76a}$$

$$\rho^2 = x^2 + y^2 + z^2, \qquad \phi = \tan^{-1}\left(\frac{y}{x}\right), \qquad \theta = \tan^{-1}\left(\frac{\sqrt{x^2 + y^2}}{z}\right) \tag{A.4.76b}$$

A.4.13.3 Cylindrical (r, ϕ, z) and Spherical (ρ, θ, ϕ)

$$r \sin\theta = \rho, \qquad \phi = \phi, \qquad r \cos\theta = z \tag{A.4.77a}$$

$$r^2 + z^2 = \rho^2, \qquad \phi = \phi, \qquad \theta = \tan^{-1}\left(\frac{r}{z}\right) \tag{A.4.77b}$$

A.4.14 PARTIAL DERIVATIVES OF THE UNIT VECTORS OF THE THREE ORTHOGONAL COORDINATE SYSTEMS

The unit vectors of a Cartesian coordinate system ($\mathbf{i}_x$, $\mathbf{i}_y$, $\mathbf{i}_z$) are oriented in a fixed direction, irrespective of the variations of any of the coordinates x, y, or z. Hence all the partial derivatives are zero, i.e.

$$\frac{\partial \mathbf{i}_x}{\partial x} = \frac{\partial \mathbf{i}_x}{\partial y} = \frac{\partial \mathbf{i}_x}{\partial z} = \frac{\partial \mathbf{i}_y}{\partial x} = \frac{\partial \mathbf{i}_y}{\partial y} = \frac{\partial \mathbf{i}_y}{\partial z} = \frac{\partial \mathbf{i}_z}{\partial x} = \frac{\partial \mathbf{i}_z}{\partial y} = \frac{\partial \mathbf{i}_z}{\partial z} = 0 \tag{A.4.78}$$

Let us now consider the variation of another unit vector, (say) $\mathbf{i}_r$ in two-dimensional cylindrical coordinate system (for simplicity of understanding), with respect to ϕ (see Figure A.4.6) .

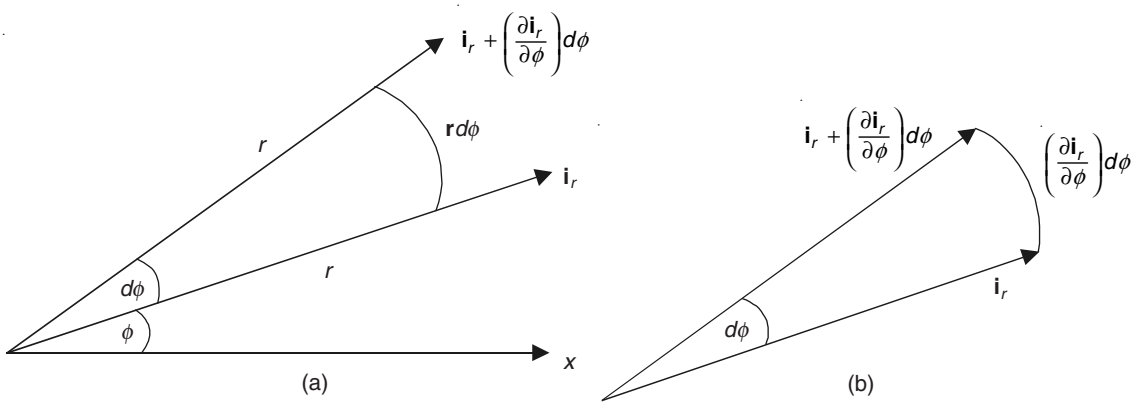


Figure A.4.6 Vector diagram for evaluating $\partial \mathbf{i}_r / \partial \phi$.

From the above figure, it should be noted that the change in $\mathbf{i}_r$ as ϕ changes is $\left(\frac{\partial \mathbf{i}_r}{\partial \phi}\right) d\phi$, and this change is in the direction of $\mathbf{i}_\phi$.

The magnitude of the partial derivative $\left(\frac{\partial \mathbf{i}_r}{\partial \phi}\right) d\phi$ divided by the magnitude of the unit vector $\mathbf{i}_r$ is $d\phi$, i.e.

$$\frac{\left|\frac{\partial \mathbf{i}_r}{\partial \phi}\right| d\phi}{\mathbf{i}_r} = d\phi \tag{A.4.79}$$

$$\therefore \quad \left|\frac{\partial \mathbf{i}_r}{\partial \phi}\right| = 1 \tag{A.4.80}$$

and its direction is that of $\mathbf{i}_\phi$.

$$\therefore \quad \frac{\partial \mathbf{i}_r}{\partial \phi} = \mathbf{i}_\phi \tag{A.4.81}$$

Hence the 9 derivatives of the cylindrical coordinate system can be shown to be:

$$\frac{\partial \mathbf{i}_r}{\partial r} = 0, \qquad \frac{\partial \mathbf{i}_\phi}{\partial r} = 0, \qquad \frac{\partial \mathbf{i}_z}{\partial r} = 0 \tag{A.4.82a}$$

$$\frac{\partial \mathbf{i}_r}{\partial \phi} = \mathbf{i}_\phi, \qquad \frac{\partial \mathbf{i}_\phi}{\partial \phi} = -\mathbf{i}_r, \qquad \frac{\partial \mathbf{i}_z}{\partial \phi} = 0 \tag{A.4.82b}$$

$$\frac{\partial \mathbf{i}_r}{\partial z} = 0, \qquad \frac{\partial \mathbf{i}_\phi}{\partial z} = 0, \qquad \frac{\partial \mathbf{i}_z}{\partial z} = 0 \tag{A.4.82c}$$

Similarly in the spherical coordinate system,

$$\frac{\partial \mathbf{i}_\rho}{\partial \rho} = 0, \qquad \frac{\partial \mathbf{i}_\theta}{\partial \rho} = 0, \qquad \frac{\partial \mathbf{i}_\phi}{\partial \rho} = 0 \tag{A.4.83a}$$

$$\frac{\partial \mathbf{i}_\rho}{\partial \theta} = \mathbf{i}_\theta, \qquad \frac{\partial \mathbf{i}_\theta}{\partial \theta} = -\mathbf{i}_\rho, \qquad \frac{\partial \mathbf{i}_\phi}{\partial \theta} = 0 \tag{A.4.83b}$$

$$\frac{\partial \mathbf{i}_\rho}{\partial \phi} = \mathbf{i}_\phi \sin\theta, \qquad \frac{\partial \mathbf{i}_\theta}{\partial \phi} = \mathbf{i}_\phi \cos\theta, \qquad \frac{\partial \mathbf{i}_\phi}{\partial \phi} = \mathbf{i}_\rho \sin\theta - \mathbf{i}_\theta \cos\theta \tag{A.4.83c}$$

A.4.15 CURL OF THE CURL OF A VECTOR

In general, in the curvilinear orthogonal coordinate system, the curl curl **A** is

$$\begin{aligned}
\nabla \times \nabla \times \mathbf{A} = {} & \mathbf{i}_1 \left(\frac{1}{h_2 h_3}\right)\left[\left(\frac{\partial}{\partial u_2}\right)\left[\left(\frac{h_3}{h_1 h_2}\right)\left\{\left(\frac{\partial}{\partial u_1}\right)(h_2 A_2) - \left(\frac{\partial}{\partial u_2}\right)(h_1 A_1)\right\}\right]\right. \\
& \left. - \left(\frac{\partial}{\partial u_3}\right)\left[\left(\frac{h_2}{h_3 h_1}\right)\left\{\left(\frac{\partial}{\partial u_3}\right)(h_1 A_1) - \left(\frac{\partial}{\partial u_1}\right)(h_3 A_3)\right\}\right]\right] \\
& + \mathbf{i}_2 \left(\frac{1}{h_3 h_1}\right)\left[\left(\frac{\partial}{\partial u_3}\right)\left[\left(\frac{h_1}{h_2 h_3}\right)\left\{\left(\frac{\partial}{\partial u_2}\right)(h_3 A_3) - \left(\frac{\partial}{\partial u_3}\right)(h_2 A_2)\right\}\right]\right. \\
& \left. - \left(\frac{\partial}{\partial u_1}\right)\left[\left(\frac{h_3}{h_1 h_2}\right)\left\{\left(\frac{\partial}{\partial u_1}\right)(h_2 A_2) - \left(\frac{\partial}{\partial u_2}\right)(h_1 A_1)\right\}\right]\right]
\end{aligned}$$

$$+\mathbf{i}_3\left(\frac{1}{h_1h_2}\right)\left[\left(\frac{\partial}{\partial u_1}\right)\left[\left(\frac{h_2}{h_3h_1}\right)\left\{\left(\frac{\partial}{\partial u_3}\right)(h_1A_1)-\left(\frac{\partial}{\partial u_1}\right)(h_3A_3)\right\}\right]\right.$$

$$\left.-\left(\frac{\partial}{\partial u_2}\right)\left[\left(\frac{h_1}{h_2h_3}\right)\left\{\left(\frac{\partial}{\partial u_2}\right)(h_3A_3)-\left(\frac{\partial}{\partial u_3}\right)(h_2A_2)\right\}\right]\right] \quad \text{(A.4.84)}$$

Or in metric coefficient form,

$$\text{curl curl } \mathbf{A} = \mathbf{i}_1\sqrt{\frac{g_{11}}{g}}\left\{\left(\frac{\partial}{\partial u_2}\right)\Gamma_3-\left(\frac{\partial}{\partial u_3}\right)\Gamma_2\right\}$$

$$+\mathbf{i}_2\sqrt{\left(\frac{g_{22}}{g}\right)}\left\{\left(\frac{\partial}{\partial u_3}\right)\Gamma_1-\left(\frac{\partial}{\partial u_1}\right)\Gamma_3\right\}$$

$$+\mathbf{i}_3\sqrt{\left(\frac{g_{33}}{g}\right)}\left\{\left(\frac{\partial}{\partial u_1}\right)\Gamma_2-\left(\frac{\partial}{\partial u_2}\right)\Gamma_1\right\}$$

where

$$\Gamma_1=\left(\frac{g_{11}}{\sqrt{g}}\right)\left[\left(\frac{\partial}{\partial u_2}\right)\left(\sqrt{g_{33}}A_3\right)-\left(\frac{\partial}{\partial u_3}\right)\left(\sqrt{g_{22}}A_2\right)\right]$$

$$\Gamma_2=\left(\frac{g_{22}}{\sqrt{g}}\right)\left[\left(\frac{\partial}{\partial u_3}\right)\left(\sqrt{g_{11}}A_1\right)-\left(\frac{\partial}{\partial u_1}\right)\left(\sqrt{g_{33}}A_3\right)\right]$$

$$\Gamma_3=\left(\frac{g_{33}}{\sqrt{g}}\right)\left[\left(\frac{\partial}{\partial u_1}\right)\left(\sqrt{g_{22}}A_2\right)-\left(\frac{\partial}{\partial u_2}\right)\left(\sqrt{g_{11}}A_1\right)\right] \quad \text{(A.4.85)}$$

In the three orthogonal coordinate systems:

Cartesian:
$$\nabla\times\nabla\times\mathbf{A}=\mathbf{i}_x\left[\frac{\partial^2A_y}{\partial x\partial y}-\frac{\partial^2A_x}{\partial y^2}-\frac{\partial^2A_x}{\partial z^2}+\frac{\partial^2A_z}{\partial z\partial x}\right]$$

$$+\mathbf{i}_y\left[\frac{\partial^2A_z}{\partial y\partial z}-\frac{\partial^2A_y}{\partial z^2}-\frac{\partial^2A_y}{\partial x^2}+\frac{\partial^2A_x}{\partial x\partial y}\right]$$

$$+\mathbf{i}_z\left[\frac{\partial^2A_x}{\partial z\partial x}-\frac{\partial^2A_z}{\partial x^2}-\frac{\partial^2A_z}{\partial y^2}+\frac{\partial^2A_y}{\partial y\partial z}\right] \quad \text{(A.4.86)}$$

Cylindrical: $\nabla\times\nabla\times\mathbf{A}=\mathbf{i}_r\left[\left(\frac{1}{r}\right)\left(\frac{\partial^2A_\phi}{\partial r\partial\phi}\right)+\left(\frac{1}{r^2}\right)\left(\frac{\partial A_\phi}{\partial\phi}\right)-\left(\frac{1}{r^2}\right)\left(\frac{\partial^2A_r}{\partial\phi^2}\right)-\left(\frac{\partial^2A_r}{\partial z^2}\right)+\left(\frac{\partial^2A_z}{\partial r\partial z}\right)\right]$

$$+\mathbf{i}_\phi\left[\left(\frac{1}{r}\right)\left(\frac{\partial^2A_z}{\partial\phi\partial z}\right)-\left(\frac{\partial^2A_\phi}{\partial z^2}\right)-\left(\frac{\partial^2A_\phi}{\partial r^2}\right)-\left(\frac{1}{r}\right)\left(\frac{\partial A_\phi}{\partial r}\right)+\frac{A_\phi}{r^2}\right.$$

$$+\left(\frac{1}{r}\right)\left(\frac{\partial^2 A_r}{\partial r\partial\phi}\right)-\left(\frac{1}{r^2}\right)\left(\frac{\partial A_r}{\partial\phi}\right)\Bigg]$$

$$+\mathbf{i}_z\left[\left(\frac{\partial^2 A_r}{\partial z\partial r}\right)+\left(\frac{1}{r}\right)\left(\frac{\partial A_r}{\partial z}\right)-\left(\frac{\partial^2 A_z}{\partial r^2}\right)-\left(\frac{1}{r}\right)\left(\frac{\partial A_z}{\partial r}\right)\right.$$

$$\left.-\left(\frac{1}{r^2}\right)\left(\frac{\partial^2 A_z}{\partial\phi^2}\right)+\left(\frac{1}{r}\right)\left(\frac{\partial^2 A_\phi}{\partial\phi\partial z}\right)\right] \qquad \text{(A.4.87)}$$

Spherical: $\nabla\times\nabla\times\mathbf{A}=\mathbf{i}_\rho\left[\left(\frac{1}{\rho}\right)\left(\frac{\partial^2 A_\theta}{\partial\rho\partial\theta}\right)+\left(\frac{\cot\theta}{\rho}\right)\left(\frac{\partial A_\theta}{\partial\rho}\right)+\left(\frac{\cot\theta}{\rho^2}\right)A_\theta+\left(\frac{1}{\rho^2}\right)\left(\frac{\partial A_\theta}{\partial\theta}\right)\right.$

$$-\left(\frac{\cot\theta}{\rho^2}\right)\left(\frac{\partial A_\rho}{\partial\theta}\right)-\left(\frac{1}{\rho^2}\right)\left(\frac{\partial^2 A_\rho}{\partial\theta^2}\right)-\left\{\left(\frac{1}{\rho^2\sin^2\theta}\right)\right\}\left(\frac{\partial^2 A_\rho}{\partial\phi^2}\right)$$

$$\left.+\left\{\left(\frac{1}{\rho\sin\theta}\right)\right\}\left(\frac{\partial^2 A_\phi}{\partial\rho\partial\phi}\right)+\left\{\left(\frac{1}{\rho^2\sin\theta}\right)\right\}\left(\frac{\partial A_\phi}{\partial\phi}\right)\right]$$

$$\mathbf{i}_\theta\left[\left\{\left(\frac{1}{\rho^2\sin\theta}\right)\right\}\left(\frac{\partial^2 A_\phi}{\partial\theta\partial\phi}\right)+\left(\frac{\cot\theta}{\rho^2\sin\theta}\right)\left(\frac{\partial A_\phi}{\partial\phi}\right)-\left\{\left(\frac{1}{\rho^2\sin^2\theta}\right)\right\}\left(\frac{\partial^2 A_\theta}{\partial\phi^2}\right)\right.$$

$$\left.-\left(\frac{2}{\rho}\right)\left(\frac{\partial A_\theta}{\partial\rho}\right)-\left(\frac{\partial^2 A_\theta}{\partial\rho^2}\right)+\left(\frac{1}{\rho}\right)\left(\frac{\partial^2 A_\rho}{\partial\theta\partial\rho}\right)\right]$$

$$\mathbf{i}_\phi\left[\left\{\left(\frac{1}{\rho\sin\theta}\right)\right\}\left(\frac{\partial^2 A_\phi}{\partial\phi\partial\rho}\right)-\left(\frac{2}{\rho}\right)\left(\frac{\partial A_\theta}{\partial\rho}\right)-\left(\frac{\partial^2 A_\phi}{\partial\rho^2}\right)+\left\{\left(\frac{1}{\rho^2\sin^2\theta}\right)\right\}A\phi\right.$$

$$-\left(\frac{\cot\theta}{\rho^2}\right)\left(\frac{\partial A_\phi}{\partial\theta}\right)-\left(\frac{1}{\rho^2}\right)\left(\frac{\partial^2 A_\phi}{\partial\theta^2}\right)+\left\{\left(\frac{1}{\rho^2\sin\theta}\right)\right\}\left(\frac{\partial^2 A_\theta}{\partial\phi\partial\theta}\right)$$

$$\left.-\left\{\left(\frac{\cot\theta}{\rho^2\sin\theta}\right)\right\}\left(\frac{\partial A_\theta}{\partial\phi}\right)\right] \qquad \text{(A.4.88)}$$

A.4.16 VECTOR LAPLACIAN (= $\nabla^2\mathbf{A}$)

The vector Laplacian, when applied to a vector is

$$\nabla^2\mathbf{A}=\text{grad div }\mathbf{A}-\text{curl curl }\mathbf{A}$$

and using the scalar operator ∇^2 on the vector, we get in the curvilinear coordinate system:

$$\nabla^2\mathbf{A}=\left(\frac{1}{h_1h_2h_3}\right)\left[\left(\frac{\partial}{\partial u_1}\right)\left\{\left(\frac{h_2h_3}{h_1}\right)\left(\frac{\partial}{\partial u_1}\right)(\mathbf{i}_1A_1+\mathbf{i}_2A_2+\mathbf{i}_3A_3)\right\}\right.$$

$$+\left(\frac{\partial}{\partial u_2}\right)\left\{\left(\frac{h_3h_1}{h_2}\right)\left(\frac{\partial}{\partial u_2}\right)(\mathbf{i}_1A_1+\mathbf{i}_2A_2+\mathbf{i}_3A_3)\right\}$$

$$+\left(\frac{\partial}{\partial u_3}\right)\left\{\left(\frac{h_1h_2}{h_3}\right)\left(\frac{\partial}{\partial u_3}\right)(\mathbf{i}_1A_1+\mathbf{i}_2A_2+\mathbf{i}_3A_3)\right\}\Bigg] \quad \text{(A.4.89)}$$

Or in terms of metric coefficients,

$$\nabla^2\mathbf{A} = \mathbf{i}_1\left[\left(\frac{1}{\sqrt{g_{11}}}\right)\left(\frac{\partial T}{\partial u_1}\right)+\left(\sqrt{\frac{g_{11}}{g}}\right)\left\{\left(\frac{\partial\Gamma_2}{\partial u_3}\right)-\left(\frac{\partial\Gamma_3}{\partial u_2}\right)\right\}\right]$$

$$+\mathbf{i}_2\left[\left(\frac{1}{\sqrt{g_{22}}}\right)\left(\frac{\partial T}{\partial u_2}\right)+\left(\sqrt{\frac{g_{22}}{g}}\right)\left\{\left(\frac{\partial\Gamma_3}{\partial u_1}\right)-\left(\frac{\partial\Gamma_1}{\partial u_3}\right)\right\}\right]$$

$$+\mathbf{i}_3\left[\left(\frac{1}{\sqrt{g_{33}}}\right)\left(\frac{\partial T}{\partial u_3}\right)+\left(\sqrt{\frac{g_{33}}{g}}\right)\left\{\left(\frac{\partial\Gamma_1}{\partial u_2}\right)-\left(\frac{\partial\Gamma_2}{\partial u_1}\right)\right\}\right] \quad \text{(A.4.90)}$$

where

$$T=\left(\frac{1}{\sqrt{g}}\right)\left[\left(\frac{\partial}{\partial u_1}\right)\left\{\left(\sqrt{\frac{g}{g_{11}}}\right)A_{11}\right\}+\left(\frac{\partial}{\partial u_2}\right)\left\{\left(\sqrt{\frac{g}{g_{22}}}\right)A_{22}\right\}+\left(\frac{\partial}{\partial u_3}\right)\left\{\left(\sqrt{\frac{g}{g_{33}}}\right)A_{33}\right\}\right]$$

and Γ_1, Γ_2, Γ_3 are as in Eq. (A.4.85).

In the Cartesian coordinates:

$$\nabla^2\mathbf{A} = \mathbf{i}_x\left(\frac{\partial^2A_x}{\partial x^2}+\frac{\partial^2A_x}{\partial y^2}+\frac{\partial^2A_x}{\partial z^2}\right)$$

$$+\mathbf{i}_y\left(\frac{\partial^2A_y}{\partial x^2}+\frac{\partial^2A_y}{\partial y^2}+\frac{\partial^2A_y}{\partial z^2}\right)$$

$$+\mathbf{i}_z\left(\frac{\partial^2A_z}{\partial x^2}+\frac{\partial^2A_z}{\partial y^2}+\frac{\partial^2A_z}{\partial z^2}\right) \quad \text{(A.4.91)}$$

In the cylindrical coordinae system:

$$\nabla^2\mathbf{A} = \mathbf{i}_r\left\{\left(\frac{\partial^2A_r}{\partial r^2}\right)+\left(\frac{1}{r}\right)\left(\frac{\partial A_r}{\partial r}\right)-\left(\frac{1}{r^2}\right)A_r+\left(\frac{1}{r^2}\right)\left(\frac{\partial^2A_r}{\partial\phi^2}\right)+\left(\frac{\partial^2A_r}{\partial z^2}\right)-\left(\frac{2}{r^2}\right)\left(\frac{\partial A_\phi}{\partial\phi}\right)\right\}$$

$$+\mathbf{i}_\phi\left\{\left(\frac{\partial^2A_\phi}{\partial r^2}\right)+\left(\frac{1}{r}\right)\left(\frac{\partial A_\phi}{\partial r}\right)-\left(\frac{1}{r^2}\right)A_\phi+\left(\frac{1}{r^2}\right)\left(\frac{\partial^2A_\phi}{\partial\phi^2}\right)+\left(\frac{\partial^2A_\phi}{\partial z^2}\right)+\left(\frac{2}{r^2}\right)\left(\frac{\partial A_r}{\partial\phi}\right)\right\}$$

$$+\mathbf{i}_z\left\{\left(\frac{\partial^2A_z}{\partial r^2}\right)+\left(\frac{1}{r}\right)\left(\frac{\partial A_z}{\partial r}\right)+\left(\frac{1}{r^2}\right)\left(\frac{\partial^2A_z}{\partial\phi^2}\right)+\left(\frac{\partial^2A_z}{\partial z^2}\right)\right\} \quad \text{(A.4.92)}$$

In the spherical coordinate system:

$$\begin{aligned}\nabla^2\mathbf{A} = \mathbf{i}_\rho &\left[\left(\frac{\partial^2 A_\rho}{\partial\rho^2}\right)+\left(\frac{2}{\rho}\right)\left(\frac{\partial A_\rho}{\partial\rho}\right)-\left(\frac{2}{\rho^2}\right)A_\rho+\left(\frac{1}{\rho^2}\right)\left(\frac{\partial^2 A_\rho}{\partial\theta^2}\right)\right.\\ &+\left(\frac{\cot\theta}{\rho^2}\right)\left(\frac{\partial A_\rho}{\partial\theta}\right)+\left(\frac{1}{\rho^2\sin^2\theta}\right)\left(\frac{\partial^2 A_\rho}{\partial\phi^2}\right)\\ &\left.-\left(\frac{2}{\rho^2}\right)\left(\frac{\partial A_\theta}{\partial\theta}\right)-\left(\frac{2\cot\theta}{\rho^2}\right)A_\theta-\left(\frac{2}{\rho^2\sin\theta}\right)\left(\frac{\partial A_\phi}{\partial\phi}\right)\right]\\ +\mathbf{i}_\theta &\left[\left(\frac{\partial^2 A_\theta}{\partial\rho^2}\right)+\left(\frac{2}{\rho}\right)\left(\frac{\partial A_\theta}{\partial\rho}\right)-\left(\frac{1}{\rho^2\sin^2\theta}\right)A_\theta\right.\\ &+\left(\frac{1}{\rho^2}\right)\left(\frac{\partial^2 A_\theta}{\partial\theta^2}\right)+\left(\frac{\cot\theta}{\rho^2}\right)\left(\frac{\partial A_\theta}{\partial\theta}\right)+\left\{\left(\frac{1}{\rho^2\sin^2\theta}\right)\right\}\left(\frac{\partial^2 A_\theta}{\partial\phi^2}\right)\\ &\left.-\left\{\left(\frac{2\cot\theta}{\rho^2\sin\theta}\right)\right\}\left(\frac{\partial A_\phi}{\partial\phi}\right)+\left(\frac{2}{\rho^2}\right)\left(\frac{\partial A_\rho}{\partial\theta}\right)\right]\\ +\mathbf{i}_\phi &\left[\left(\frac{\partial^2 A_\phi}{\partial\rho^2}\right)+\left(\frac{2}{\rho}\right)\left(\frac{\partial A_\phi}{\partial\rho}\right)-\left(\frac{1}{\rho^2\sin^2\theta}\right)A_\phi\right.\\ &+\left(\frac{1}{\rho^2}\right)\left(\frac{\partial^2 A_\phi}{\partial\theta^2}\right)+\left(\frac{\cot\theta}{\rho^2}\right)\left(\frac{\partial A_\phi}{\partial\theta}\right)+\left\{\left(\frac{1}{\rho^2\sin^2\theta}\right)\right\}\left(\frac{\partial^2 A_\phi}{\partial\phi^2}\right)\\ &\left.+\left\{\left(\frac{2\cot\theta}{\rho^2\sin\theta}\right)\right\}\left(\frac{\partial A_\theta}{\partial\phi}\right)+\left(\frac{2}{\rho^2\sin\theta}\right)\left(\frac{\partial A_\rho}{\partial\phi}\right)\right]\end{aligned}\tag{A.4.93}$$

A.4.17 GRADIENT OF THE DIVERGENCE OF A VECTOR

In the curvilinear coordinate system:

$$\begin{aligned}\text{grad div }\mathbf{A} = \nabla(\nabla\cdot\mathbf{A}) = \mathbf{i}_1 &\left(\frac{1}{h_1}\right)\left(\frac{\partial}{\partial u_1}\right)\left[\left(\frac{1}{h_1h_2h_3}\right)\left\{\left(\frac{\partial}{\partial u_1}\right)(h_2h_3A_1)\right.\right.\\ &\left.\left.+\left(\frac{\partial}{\partial u_2}\right)(h_3h_1A_2)+\left(\frac{\partial}{\partial u_3}\right)(h_1h_2A_3)\right\}\right]\\ +\mathbf{i}_2 &\left(\frac{1}{h_2}\right)\left(\frac{\partial}{\partial u_2}\right)\left[\left(\frac{1}{h_1h_2h_3}\right)\left\{\left(\frac{\partial}{\partial u_1}\right)(h_2h_3A_1)\right.\right.\\ &\left.\left.+\left(\frac{\partial}{\partial u_2}\right)(h_3h_1A_2)+\left(\frac{\partial}{\partial u_3}\right)(h_1h_2A_3)\right\}\right]\end{aligned}$$

$$+\mathbf{i}_3\left(\frac{1}{h_3}\right)\left(\frac{\partial}{\partial u_3}\right)\left[\left(\frac{1}{h_1h_2h_3}\right)\left\{\left(\frac{\partial}{\partial u_1}\right)(h_2h_3A_1)+\left(\frac{\partial}{\partial u_2}\right)(h_3h_1A_2)+\left(\frac{\partial}{\partial u_3}\right)(h_1h_2A_3)\right\}\right] \tag{A.4.94}$$

In terms of the metric coefficients,

$$\text{grad div } \mathbf{A} = \nabla(\nabla\cdot\mathbf{A}) = \left(\frac{\mathbf{i}_1}{\sqrt{g_{11}}}\right)\left(\frac{\partial T}{\partial u_1}\right)+\left(\frac{\mathbf{i}_2}{\sqrt{g_{22}}}\right)\left(\frac{\partial T}{\partial u_2}\right)+\left(\frac{\mathbf{i}_3}{\sqrt{g_{33}}}\right)\left(\frac{\partial T}{\partial u}\right) \tag{A.4.95a}$$

where

$$T = \left(\frac{1}{\sqrt{g}}\right)\left[\frac{\partial}{\partial u_1}\left\{\left(\sqrt{\frac{g}{g_{11}}}\right)A_1\right\}+\frac{\partial}{\partial u_2}\left\{\left(\sqrt{\frac{g}{g_{22}}}\right)A_2\right\}+\frac{\partial}{\partial u_3}\left\{\left(\sqrt{\frac{g}{g_{33}}}\right)A_3\right\}\right] \tag{A.4.95b}$$

In the three orthogonal coordinate systems:

Cartesian:

$$\begin{aligned}\nabla(\nabla\cdot\mathbf{A}) = {} & \mathbf{i}_x\left(\frac{\partial^2A_x}{\partial x^2}+\frac{\partial^2A_y}{\partial x\partial y}+\frac{\partial^2A_z}{\partial x\partial z}\right)\\ & +\mathbf{i}_y\left(\frac{\partial^2A_x}{\partial x\partial y}+\frac{\partial^2A_y}{\partial y^2}+\frac{\partial^2A_z}{\partial y\partial z}\right)\\ & +\mathbf{i}_z\left(\frac{\partial^2A_x}{\partial x\partial z}+\frac{\partial^2A_y}{\partial y\partial z}+\frac{\partial^2A_z}{\partial z^2}\right)\end{aligned} \tag{A.4.96}$$

Cylindrical:

$$\begin{aligned}\nabla(\nabla\cdot\mathbf{A}) = {} & \mathbf{i}_r\left\{\left(\frac{\partial^2A_r}{\partial r^2}\right)+\left(\frac{1}{r}\right)\left(\frac{\partial A_r}{\partial r}\right)-\left(\frac{A_r}{r^2}\right)\right.\\ & \left.+\left(\frac{1}{r}\right)\left(\frac{\partial^2A_\phi}{\partial r\partial\phi}\right)-\left(\frac{1}{r^2}\right)\left(\frac{\partial A_\phi}{\partial\phi}\right)+\left(\frac{\partial^2A_z}{\partial r\partial z}\right)\right\}\\ & +\mathbf{i}_\phi\left\{\left(\frac{1}{r}\right)\left(\frac{\partial^2A_r}{\partial r\partial\phi}\right)+\left(\frac{1}{r^2}\right)\left(\frac{\partial A_r}{\partial\phi}\right)\right.\\ & \left.+\left(\frac{1}{r^2}\right)\left(\frac{\partial^2A_\phi}{\partial\phi^2}\right)+\left(\frac{1}{r}\right)\left(\frac{\partial^2A_z}{\partial\phi\partial z}\right)\right\}\\ & +\mathbf{i}_z\left\{\left(\frac{\partial^2A_r}{\partial r\,\partial z}\right)+\left(\frac{1}{r}\right)\left(\frac{\partial A_r}{\partial z}\right)+\left(\frac{1}{r}\right)\left(\frac{\partial^2A_\phi}{\partial\phi\partial z}\right)+\left(\frac{\partial^2A_z}{\partial z^2}\right)\right\}\end{aligned} \tag{A.4.97}$$

$$\text{Spherical: } \nabla(\nabla\cdot\mathbf{A}) = \mathbf{i}_\rho\left[\left(\frac{\partial^2 A_\rho}{\partial\rho^2}\right)+\left(\frac{2}{\rho}\right)\left(\frac{\partial A_\rho}{\partial\rho}\right)-\left(\frac{2}{\rho^2}\right)A_\rho\right.$$
$$+\left(\frac{\cot\theta}{\rho^2}\right)\left(\frac{\partial A_\theta}{\partial\rho}\right)+\left(\frac{1}{\rho}\right)\left(\frac{\partial^2 A_\theta}{\partial\rho\partial\theta}\right)-\left(\frac{1}{\rho^2}\right)\left(\frac{\partial A_\theta}{\partial\theta}\right)$$
$$\left.-\left(\frac{\cot\theta}{\rho^2}\right)A_\theta+\left(\frac{1}{\rho\sin\theta}\right)\left(\frac{\partial^2 A_\phi}{\partial\phi\partial\rho}\right)-\left(\frac{1}{\rho^2\sin\theta}\right)\left(\frac{\partial A_\phi}{\partial\phi}\right)\right]$$
$$+\mathbf{i}_\theta\left[\left(\frac{1}{\rho}\right)\left(\frac{\partial^2 A_\rho}{\partial\rho\partial\theta}\right)+\left(\frac{2}{\rho^2}\right)\left(\frac{\partial A_\rho}{\partial\theta}\right)-\left(\frac{1}{\rho^2\sin^2\theta}\right)A_\theta\right.$$
$$+\left(\frac{1}{\rho^2}\right)\left(\frac{\partial^2 A_\theta}{\partial\theta^2}\right)+\left(\frac{\cot\theta}{\rho^2}\right)\left(\frac{\partial A_\theta}{\partial\theta}\right)$$
$$\left.+\left(\frac{1}{\rho^2\sin\theta}\right)\left(\frac{\partial^2 A_\phi}{\partial\theta\partial\phi}\right)-\left(\frac{\cot\theta}{\rho^2\sin\theta}\right)\left(\frac{\partial A_\phi}{\partial\phi}\right)\right]$$
$$+\mathbf{i}_\phi\left[\left(\frac{1}{\rho\sin\theta}\right)\left(\frac{\partial^2 A_\rho}{\partial\rho\partial\phi}\right)+\left(\frac{2}{\rho^2\sin\theta}\right)\left(\frac{\partial A_\rho}{\partial\phi}\right)\right.$$
$$\left.+\left(\frac{\cot\theta}{\rho^2\sin\theta}\right)\left(\frac{\partial A_\theta}{\partial\phi}\right)+\left(\frac{1}{\rho^2\sin\theta}\right)\left(\frac{\partial^2 A_\theta}{\partial\theta\partial\phi}\right)+\left(\frac{1}{\rho^2\sin^2\theta}\right)\left(\frac{\partial^2 A_\phi}{\partial\phi^2}\right)\right] \tag{A.4.98}$$

A.4.18 DIVERGENCE OF THE CURL OF A VECTOR

This is a vector identity, which in curvilinear coordinates is:

$$\nabla\cdot(\nabla\times\mathbf{A}) = \left(\frac{1}{h_1h_2h_3}\right)\left[\left(\frac{\partial^2}{\partial u_1\partial u_2}\right)(h_3A_3)-\left(\frac{\partial^2}{\partial u_3\partial u_1}\right)(h_2A_2)+\left(\frac{\partial^2}{\partial u_2\partial u_3}\right)(h_1A_1)\right.$$
$$\left.-\left(\frac{\partial^2}{\partial u_2\partial u_1}\right)(h_3A_3)+\left(\frac{\partial^2}{\partial u_1\partial u_3}\right)(h_2A_2)-\left(\frac{\partial^2}{\partial u_3\partial u_2}\right)(h_1A_1)\right]$$
$$= 0 \tag{A.4.99}$$

(Note that the order of differentiation is immaterial.)

A.4.19 CURL OF THE GRADIENT OF A SCALAR

This is another vector identity, i.e.

$$\nabla\times(\nabla\varphi) = \left(\frac{1}{h_1h_2h_3}\right)\left[\mathbf{i}_1h_1\left(\frac{\partial^2\varphi}{\partial u_2\partial u_3}-\frac{\partial^2\varphi}{\partial u_3\partial u_2}\right)\right.$$
$$\left.+\mathbf{i}_2h_2\left(\frac{\partial^2\varphi}{\partial u_3\partial u_1}-\frac{\partial^2\varphi}{\partial u_1\partial u_3}\right)+\mathbf{i}_3h_3\left(\frac{\partial^2\varphi}{\partial u_1\partial u_2}-\frac{\partial^2\varphi}{\partial u_2\partial u_1}\right)\right]$$
$$= 0 \tag{A.4.100}$$

Appendix 5

Some Important Physical Constants

Permittivity of free space

$$\varepsilon_0 = 8.854 \times 10^{-12} \left(\approx \frac{10^{-9}}{36\pi} \right) \text{ farads/metre}$$

Permeability of free space (absolute permeability)

$$\mu_0 = 4\pi \times 10^{-7} \text{ henry/metre}$$

Velocity of light in vacuum, $c_0 = \dfrac{1}{\sqrt{\mu_0 \varepsilon_0}}$

$$= 2.998 \times 10^8 \ (\approx 3 \times 10^8) \text{ metres/sec}$$

Intrinsic impedance of vacuum, $Z_0 = \sqrt{\dfrac{\mu_0}{\varepsilon_0}}$

$$= 376.7 \ (\approx 120\pi) \ \Omega$$

Charge of an electron, $e = 1.60202 \times 10^{-19}$ coulomb

Electron rest mass, $m_e = 9.1083 \times 10^{-31}$ kg = 0.511 MeV

(Charge/mass) of electron $(e/m_e) = 1.7589 \times 10^{11}$ coulombs/kg

Proton rest mass, $m_p = 1.6724 \times 10^{-27}$ kg = 938 MeV

Classical electron radius, $\dfrac{e^2}{4\pi^2 \varepsilon_0 m_e c^2} = 2.8178 \times 10^{-15}$ metres

Bohr radius, $\dfrac{4\pi\varepsilon_0 h^2}{4\pi^2 m_e e^2} = 5.2917 \times 10^{-11}$ metres

Planck's constant, $h = 6.6252 \times 10^{-34}$ joules-sec

Boltzmann's constant, $k = 1.3804 \times 10^{-23}$ joules/K

Avogadro's number, $N_A = 6.0249 \times 10^{26}$ /kg

Appendix 6

A Circuit for Experimental Verification of One of the Maxwell's Equations (Faraday's Law of Induction)

A.6.1 INTRODUCTION

The subject-matter discussed here is based on the material contained in Chapters 6, 12, 13, 15, and 17.

The basic elements in circuit theory are two-terminal generators, resistors, capacitors, and inductors (coils). Any number of these elements, connected by conducting wires, makes an electrical network. From the approach of field theory, the analysis of such a network would be extremely complicated because the spatial distribution of the elements and the orientation of the connecting conducting wires can be geometrically highly irregular. The boundaries would have no simple regular shapes and the medium would be in-homogeneous. In most of the cases no closed-form analytical solution would be possible.

Yet very large number of such networks have been and are being solved by the methods of circuit theory; and the solutions thus obtained have been the basis of the designs of many practical devices which we encounter everyday. The agreement between the practice and the predictions based on this simple theory has been excellent. That the circuit theory is an approximate theory is very obvious, and yet the accuracy of its results in a very large number of cases has been so high, that no better solution is required. As with any approximate theory, the circuit theory is valid, only under certain circumstances. A knowledge of the range of its validity is important for the engineers dealing with it.

A.6.2 THE CIRCUIT THEORY AND ITS LIMITATIONS

The circuit theory is essentially a one-dimensional simplification of the field theory which deals with two- and three-dimensional spatial distributions. It (circuit theory) cannot consider the effects of space boundary conditions as is done in field theory problems.

The circuit theory is based on the two following assumptions:

1. No charges are accumulated in the regions of the nodes of a circuit and on the interconnecting conductors. This condition then permits the validity of the Kirchhoff's first law to be extended to time-varying currents. This law is effectively a special case of the equation of continuity in integral form, i.e.

$$\oiint_S \mathbf{J} \cdot d\mathbf{S} = 0 \tag{A.6.1}$$

For circuits, it is expressed as

$$\sum_{\text{any node}} i(t) = 0 \tag{A.6.2}$$

which is an extension of Eq. (6.26), now taking account of the time-varying currents as well.

2. Kirchhoff's second law is for the potential differences in a circuit. It can be stated as "the line-integral of the 'total' electric field intensity and can now be associated uniquely with any pairs of nodes of two-terminal network elements from which the network is built". Under this condition, the mathematical statement of the law can be written in a slightly different form from Eq. (6.38), as

$$\sum \mathscr{E}(t) - \sum [V_i(t) - V_j(t)] = 0 \qquad \text{(for any closed loop)} \tag{A.6.3}$$

where the difference $[V_i(t) - V_j(t)]$ is the potential difference between the terminals i and j of two-terminal elements along a given closed contour.

Both, this equation and Eq. (6.38) are effectively special cases of the integral form of the Maxwell's equation for Faraday's induction law, i.e.

$$\oint_C \mathbf{E} \cdot d\mathbf{l} = 0 \tag{A.6.4}$$

When only resistors, capacitors, and inductors are present in the circuit, Eq. (A.6.3) takes the form

$$\sum \mathscr{E}(t) - \sum \left[Ri(t) + L\frac{di(t)}{dt} + \frac{1}{C}\int i(t)dt \right] = 0 \tag{A.6.5}$$

When the source emf produces sinusoidal time-varying currents, by using the complex notation, the above equation becomes

$$\sum V - \sum \left[RI_R + j\omega L I_L + \frac{1}{j\omega C} I_C \right] = 0 \tag{A.6.6}$$

where V's and I's represent complex amplitudes.

There are two limitations implicitely assumed in these two laws. Since it was assumed that no charges are present on interconnecting conductors, it implies that the electric field is localized inside the elements themselves. This means that the displacement current density ($= \partial D/\partial t$) is negligible, and hence the retardation effects are ignored. Hence for sinusoidally time-varying currents of frequency $f(= \omega/2\pi)$, the largest dimension of the circuit (= d, say) must be $<< 1/\omega\sqrt{\mu\varepsilon}$.

So it could be seen that the Kirchhoff's two laws would enable us to analyze such circuits or networks to evaluate the branch currents and the potential differences (P.D's.) between any

two nodes uniquely. This means that in such a network, the P.D. between any two points (or nodes) depends only on the potential values of the nodes and is independent of the path traversed by the contour, i.e. for the two nodes A and B on a contour of the network,

$$V_{AB} = -\int_{A}^{B} \mathbf{E} \cdot d\mathbf{l} = V_A - V_B \tag{A.6.7}$$

A.6.3 CIRCULATING CURRENT CIRCUIT

This (the above) is in general true for circuits in which the energy source is a battery or ac generator (or oscillator). But there are circuits in which the time-varying currents can be produced by, say, a time-varying magnetic field distributed in a region which is enclosed by a contour along which a circuit containing passive elements can be constructed. In such a circuit, it will be found that the P.D. between the two nodes (of an element) would not be a unique function of the potentials of the concerned nodes, but would also depend on the orientation and location of the contour along which the distance between the nodes is being traversed. What is being implied is that up to Chapter 6 and in the present discussion (in this Appendix so far), from Eq. (A.6.4), the relationship of the electric intensity vector **E** with the potential V is given by the equation

$$\mathbf{E} = -\text{ grad } V = -\nabla V \tag{A.6.8}$$

which holds true for both steady currents (time-invariant) as well as for time-varying currents (which includes sinusoidally time-varying currents as well). However we will now consider a circuit whose behaviour cannot be fully explained by the equations considered so far, and can only be explained rigorously by the integral form of Maxwell's equation for Faraday's law of induction. This circuit was first discussed by Haus and Melcher[1] and then by Woodson and Melcher[2].

The circuit under consideration consists of two resistors R_a and R_b (where $R_a \neq R_b$) looped round one limb of a rectangular laminated core, around the other limb of which is wound an exciting coil supplied from an alternating current source which in this case is a single-phase transformer (Figure A.6.1). The current in the exciting coil produces an alternating magnetic flax in the iron core, which links the loop of the circuit PR_aQR_bP, as shown in the Figure A.6.2. Since the core is made up of highly permeable iron laminations (CRGO steel), the alternating magnetic flux in it (the core), generated by the alternating current [$= i \exp(j\omega t)$] in the exciting coil, would be inside the iron laminations and there would be no leakage flux in the surrounding air-space. As the loop circuit is enclosing an alternating magnetic flux [of density $B \exp(j\omega t)$] in the iron core, there would be a 'circulating current' in it, even though there is no emf or alternating current source connected to it. The current and the induced emf in the circuit is given by the integral form of Maxwell's equation for Faraday's law of induction, i.e.

$$\oint_{C} \mathbf{E} \cdot d\mathbf{l} = -\frac{d}{dt} \iint \mathbf{B} \cdot d\mathbf{S} \tag{A.6.9}$$

where C is the contour enclosing the surface S.

In the present situation, C is the loop PR_aQR_bP and S is the area enclosed by the contour of this loop. Since all the magnetic flux is in the iron core and there is no leakage flux in the surrounding air-space, the effective value of S would be the cross-sectional area of the rectangular core ($= A$, say) where the average flux density is $B \exp(j\omega t)$ produced by an N-turn coil

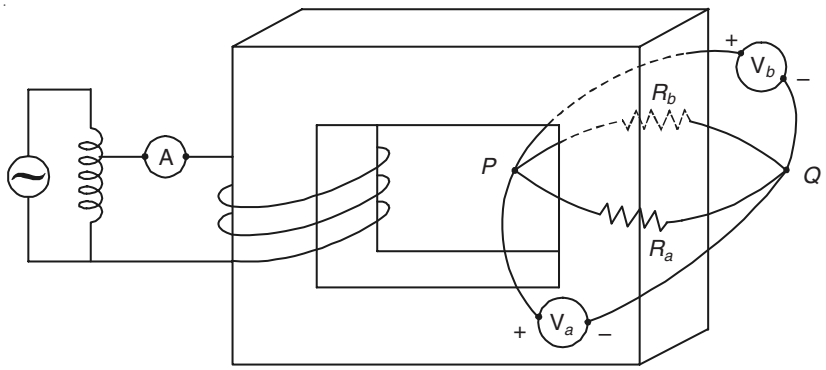


Figure A.6.1 Pictorial view of a magnetic circuit for generating circulating current.

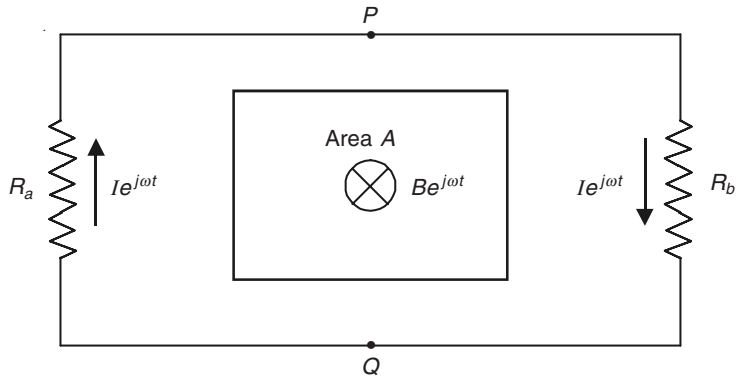


Figure A.6.2 Cross-section of the magnetic flux area within the circulating current circuit.

enclosing the other limb of the core, and carrying a current $i \exp(j\omega t)$.

∴ The total flux linked by the loop C $(PR_aQR_bP) = \Phi = AB \exp(j\omega t)$

∴ The emf in the contour C due to the alternating flux linked by it

$$= V = \oint_C \mathbf{E} \cdot d\mathbf{l} = (R_a + R_b)\, I \exp(j\omega t) = -\frac{d}{dt}\left[AB \exp(j\omega t)\right] \tag{A.6.10}$$

where $I \exp(j\omega t)$ is the current flowing in the loop due to the induced emf (Figure A.6.2).

To find the P.D. between the points P and Q—the joints of the resistors R_a and R_b, we connect two voltmeters to these two points and position them as indicated in Figure A.6.3, i.e. V_a and V_b. To obtain the values V_a and V_b (i.e. the voltmeter readings), we need to consider the line integrals of $\mathbf{E}$ along the contours C_a and C_b as shown in Figure A.6.3. Since these two contours enclose no magnetic flux at all, we get

$$\oint_{C_a} \mathbf{E} \cdot d\mathbf{l} = V_a \exp(j\omega t) + R_a I \exp(j\omega t) = 0 \tag{A.6.11a}$$

$$\oint_{C_b} \mathbf{E} \cdot d\mathbf{l} = V_b \exp(j\omega t) - R_b I \exp(j\omega t) = 0 \tag{A.6.11b}$$

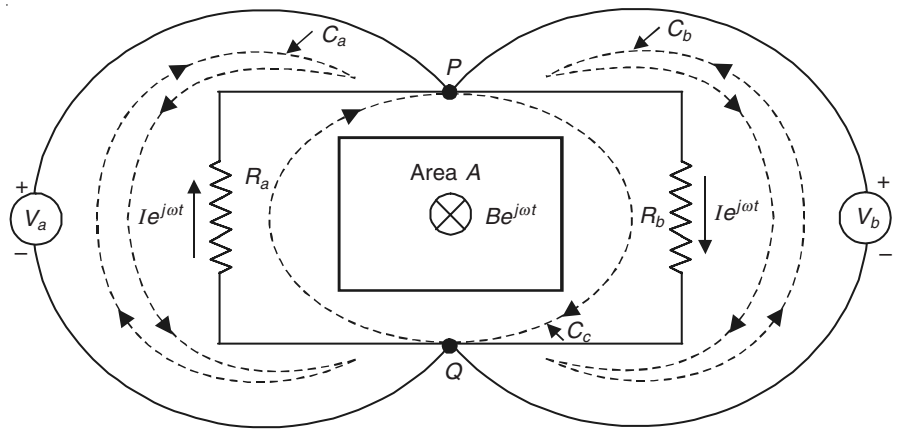


Figure A.6.3 Circulating current circuit with measuring voltmeters, and showing the contours for integration.

From Eq. (A.6.10),

$$I = -\frac{AB\omega}{R_a + R_b} \tag{A.6.12}$$

$$\therefore \quad V_a = j\frac{\omega R_a AB}{R_a + R_b} \quad \text{and} \quad V_b = -j\frac{\omega R_b AB}{R_a + R_b} \tag{A6.13}$$

Since $R_a \neq R_b$, $|V_a| \neq |V_b|$.

Thus it is seen that V_{PQ} measured by two voltmeters (of identical nature), both connected to the points P and Q but following different contours (and hence following different contours for the line-integrals as well) C_a and C_b (Figure A.6.3), does not produce a unique value. We obtain two values V_a and V_b which are unequal. An experimental verification of this double-valued nature of the P.D. is shown in Plates 4 and 5 which show the actual experimental set-up using a rectangular laminated core, and the two resistors are $R_a = 5\ \Omega$ and $R_b = 10\ \Omega$. The two voltmeters V_a and V_b have been replaced by the two beams of a double-beam CRO in which both beams are supplied from the points P and Q but the connecting leads follow the different contours C_a and C_b. From Plate 5, it is seen that the magnitudes of the two 'P.D's' are such that $V_a = {}^1/_2 V_b$. So it is obvious that in a circuit as this, the 'P.D.' between any two points is very much a function of the contour along which it is measured, and does not satisfy Eqs. (A.6.7) and (A.6.8).

Since the current in the circuit is produced by a time-varying magnetic field (distributed in space), the relationship between the electric field intensity $\mathbf{E}$ and the potential V is given by

$$\mathbf{E} = -\nabla V - \frac{\partial \mathbf{A}}{\partial t} \tag{13.49}$$

where $\mathbf{A}$ is the magnetic vector potential of the magnetic field, as we have seen earlier in Chapter 13. This equation defines the 'total electric field intensity $\mathbf{E}$', and the time-integral of the total electric field intensity is referred to as the "voltage between the two points P and Q" (for want of a better nomenclature). We designate it as

$$V_{12} = \int_P^Q \mathbf{E} \cdot d\mathbf{l} = -\int_P^Q \left(\nabla V + \frac{\partial \mathbf{A}}{\partial t}\right) \cdot d\mathbf{l} \tag{A.6.14}$$

The first term on the right-hand side is precisely the 'potential difference' between the two points, i.e.

$$V_{12} = (V_P - V_Q) - \frac{\partial}{\partial t} \int \mathbf{A} \cdot d\mathbf{l} \tag{A.6.15}$$

Hence the concepts of 'voltage between two points' and 'potential difference between two points' are different and not equal. Only in the static case, they become identical. The P.D. $(V_P - V_Q)$ is a function of the two points P and Q only. But the second term on the right-hand side is a function of the particular path between the points P and Q. Thus in the circuit of the Figures A.6.2 and A.6.3,

$$\int_{P \text{ along } R_b}^{Q} \mathbf{A} \cdot \mathbf{d}l + \int_{Q \text{ along } R_a}^{P} \mathbf{A} \cdot d\mathbf{l} = \iint \mathbf{B} \cdot d\mathbf{S} = \Phi$$

so that

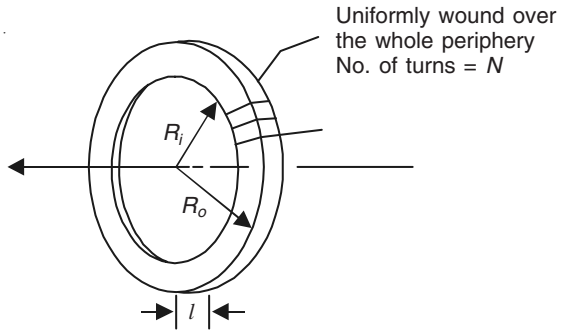
$$\int_{P \text{ along } R_b}^{Q} \mathbf{A} \cdot d\mathbf{l} = \int_{P \text{ along } R_a}^{Q} \mathbf{A} \cdot d\mathbf{l} + \Phi \tag{A.6.16}$$

$$\therefore \quad V_{12 \text{ along } R_b} = V_{12 \text{ along } R_a} - \frac{\partial \Phi}{\partial t} \tag{A.6.17}$$

where Φ is the magnetic flux through the contour PR_aQR_bP. This means that for such circuits, the voltmeter readings for measuring the voltage between two points on the circuit, would be different for different shapes of the connecting leads. This is an important conclusion for the time-varying electromagnetic fields. This conclusion violates the basic assumption of the circuit theory that the voltage between the two points in a circuit is unique because the magnetic field due to the circuit elements is concentrated inside the elements and can be ignored elsewhere. This assumption is thus seen to be incorrect when the current in the circuit is produced by a space distributed time-varying electromagnetic field.

This type of current, commonly known as circulating current (Pramanik[3]), is seen in many practical situations, such as the rotor bar currents in the short-circuited cage of squirrel cage induction motors, currents in the secondary windings of transformers under short-circuited conditions, alternating currents in the shafts of either induction or synchronous (turbo-generators and hydrogenerators) machines, arising out of magnetic dissymmetry in their cores. Other examples are the more commonly known eddy current loops in solid structures of metals (or conducting media) in the immediate vicinity of any generated alternating magnetic fields in different devices. Even the currents for cooking generated in the microwave cookers are examples of circulating currents.

Measurement of such currents poses some problems. When there are distributed multiple loops of currents in solid structures, measuring the total current in the structure is a near impossibility and it is only the surface current density which can be measured by eithor J-probes or current-density probes. Where the total current is measurable (i.e. in the shaft of synchronous machines or rotor bars of induction machines), it is preferable to use a non-contact, non-invasive device like Rogowski coil which is essentially a toroidal search coil, fixed around the current-carrying conductor (Figure A.6.4).



$$\text{Induced emf} = -N\frac{d\phi}{dt}$$

$$= N\frac{d}{dt}\iint \mathbf{B}\cdot d\mathbf{S}$$

$$= -N\frac{d}{dt}\left(\frac{\mu_0 I}{2\pi}\, l\cdot \ln\frac{R_o}{R_i}\, e^{j\omega t}\right)$$

$$= -j\frac{N\omega\mu_0 I}{2\pi}\, l\cdot \ln\frac{R_o}{R_i}\, e^{j\omega t} = V \text{ (say)}$$

$$I = \frac{2\pi V}{\mu_0 \omega N l\cdot \ln\left(\frac{R_o}{R_i}\right)}$$

Figure A.6.4 Toroidal search coil (or Rogowski coil) for non-contact measurement of current.

A.6.4 NON-CONSERVATIVE FIELDS: SOME FURTHER THOUGHTS (Later Addition)

It is obvious that in the problem we have been discussing in the previous section (i.e. Section A.6.3), the circuit enclosing the region where the time-varying magnetic field ($Be^{j\omega t}$) is distributed spatially (Figures A.6.2 and A.6.3) is a region of non-conservative field, because in this region

$$\int_C \mathbf{E}\cdot d\mathbf{l} \neq 0 \qquad \left(\text{i.e.} = -\frac{d}{dt}\iint \mathbf{B}\cdot d\mathbf{S}\right) \tag{A.6.9a}$$

It has been shown that the 'voltage drop' between any two points on a circuit along the contour enclosing the region of time-varying magnetic field is not uniquely defined. This has been shown both by mathematical analysis (based on the physics of the problem) and by experimental measurements. However there has been one idealization assumption implicit in that analysis [i.e. Eqs. (A.6.11a) and (A.6.11b), and Figure A.6.3]. It is that the two voltmeters V_a and V_b are assumed to have infinite internal resistance (i.e. $R_i \to \infty$) so that there is no current flow through them. But in real-life the internal resistance of the voltmeters would be quite large compared with the circuit resistances (i.e. in this case R_a and R_b), still R_i would have a finite value and there would be a current-flow through these two voltmeters. Hence to be rigorous, we should consider current-flow in all the three loops under consideration in Figure A.6.3. But this does not mean that the conclusions based on that analysis are incorrect. In fact we shall show that the conclusions based on this rigorous analysis also support the earlier conclusions. Furthermore the present analysis would give us a deeper insight into the behaviour of such non-conservative field regions.

Hence we will re-analysis the circuit of Figure A.6.3 and we shall consider the loop currents in the contours of complete circuit. Since we are now considering the loop currents in the

three loops, it is necessary that all the three loop currents have the same sense (i.e. in this case all the three currents be shown clockwise) and the same diagram is redrawn as Figure A.6.5. This was not an essential requirement as the currents through the voltmeters were assumed to be zero.

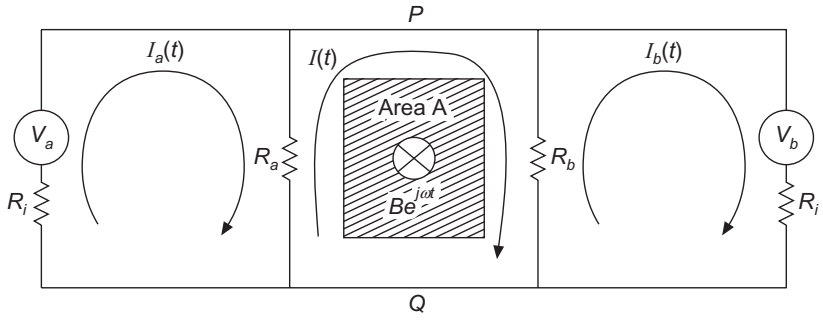


Figure A.6.5 Circulating current circuit showing the loop currents.

Since the magnetic field is time-varying, we are considering an instant of time t, when the sense of the current in each loop is the same. The central loop has two resistors R_a and R_b, whereas the voltmeters in the two side loops have the same internal resistance R_i. The loop currents are as shown in the figure. Of the three closed loops in the figure, in the two extreme loops (i.e. the one on the left and one on the right), the Kirchhoff's second law can be applied as

$$\oint_C \mathbf{E} \cdot d\mathbf{l} = 0 \tag{A.6.4}$$

whereas for the central loop which physically encloses the region where the time-varying magnetic field exists, the Faraday's law of electromagnetic induction has to be used, i.e.

$$\oint_C \mathbf{E} \cdot d\mathbf{I} = \frac{d}{dt} \iint \mathbf{B} \cdot d\mathbf{S} = \mathscr{E} \tag{A.6.9a}$$

The results would be:

$$\text{Left loop:} \quad I_a R_i + (I_a - I) R_a = 0 \tag{A.6.18}$$

$$\text{Right loop:} \quad I_b R_i + (I_b - I) R_b = 0 \tag{A.6.19}$$

$$\text{Central loop:} \quad (I - I_a) R_a + (I - I_b) R_b = \mathscr{E} \tag{A.6.20}$$

Now $R_i >> R_a$ and $R_i >> R_b$, and hence $I_a << I$ and $I_b << I_b$. Thus Eqs. (A.6.18)–A.6.20) simplify to

$$I_a R_i - I R_a \simeq 0 \tag{A.6.21}$$

$$I_b R_i - I R_b \simeq 0 \tag{A.6.22}$$

$$I (R_a + R_b) \simeq \mathscr{E} \tag{A.6.23}$$

Since each of the two voltmeters (i.e. V_a and V_b) would show reading proportional to the current passing through it, then

$$V_a = I_a R_i \simeq I R_a \tag{A.6.24}$$

and

$$V_b = I_b R_i \simeq I R_b \tag{A.6.25}$$

If the positive side of both the voltmeters is connected to the Q-side of the circuit and the negative side with the P-side, then the recorded values of V_a and V_b will be opposite in sign (as is confirmed in Plate 5 showing the V_a and V_b traces from the experiment). It should be noted that

$$\frac{V_a}{V_b} = \frac{R_a}{R_b} \text{ (and independent of the current through the control circuit)} \qquad \text{(A.6.26)}$$

$$\text{and is} \neq 1 \qquad (\because R_a \neq R_b)$$

Thus the 'voltage' between the two points P and Q is different and depends on the contour along which it is measured.

This is an important property of the non-conservative fields. Thus if we traverse on the central loop from Q to P (Q having a higher potential than P) through the resistor R_a, then the P.D. between Q and $P = (V_Q - V_P) \simeq I R_a$ as registered by the voltmeter on the left ($= V_a$). Then we complete our journey by traversing through the resistor R_b and now the P.D. between P and $Q \simeq I R_b$. So we have traversed through the whole loop in the centre, starting from and ending at, and we find that

$$V_Q - V_Q \neq 0 \qquad (\because R_a \neq R_b) \qquad \text{(A.6.27)}$$

This curious feature of non-conservative fields can be further exemplified by studying the following two circuits which are simple extrapolations of the circuit which we have been analysing so far.

In Figure A.6.6, two identical voltmeters with the same internal resistance R_i, have been apparently (?) short-circuited with the short-circuiting wire passing through the point Q as in the figure. The voltmeter V_c would show 'zero deflection', but what would be the reading of the voltmeter V_d?

In Figure A.6.7, what would be the reading of the voltmeter V_b compared to V_a with the connections to the points P and Q as shown?

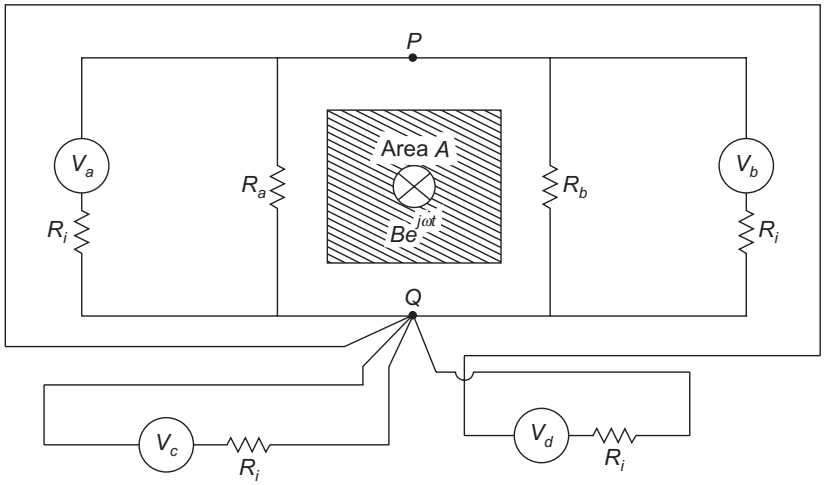


Figure A.6.6 Circulating current circuit with two additional voltmeters apparently shorted at the point Q.

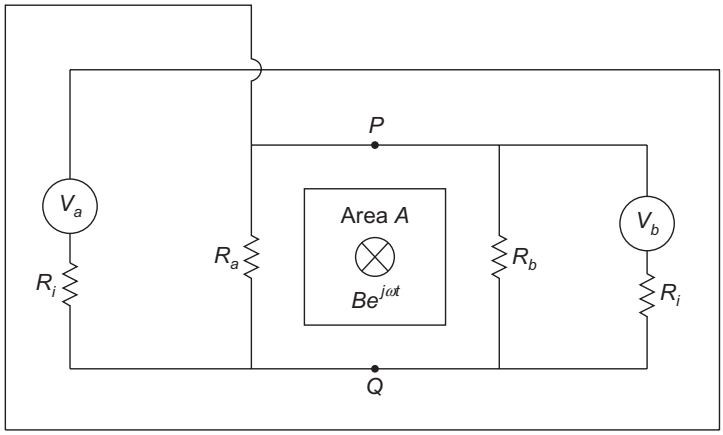


Figure A.6.7 Circulating current circuit with the voltmeter V_a connected with an additional loop around the non-conservative field region.

A.6.5 SELF-INDUCTANCE AND NON-CONSERVATIVE FIELDS

The concept of self-inductance has been discussed in reasonable detail in Sections 10.6 to 10.6.3, Sections 11.15 to 11.15.1, as well as Sections 13.4 to 13.4.6 from different approaches, i.e. Kirchhoff's Law, Law of Induction, Energy considerations as well as based on the magnetic vector potential. However now we shall have a relook at the concept from the consideration of non-conservative fields.

So we start with time-varying magnetic fields where the closed line-integral of the associated (time-varying) electric field is 'no longer' zero, i.e. for any open surface,

$$\oint \mathbf{E} \cdot d\mathbf{l} = -\frac{d}{dt} \iint \mathrm{B} \cdot d\mathrm{S} \tag{A.6.28}$$

Now, it is known that in any circuit which carries time-varying currents will have associated time-varying magnetic fields which would produce induced electric fields. So even a simple circuit with such currents would have to take account of these effects (i.e. account for the inductance of the circuit).

The main point to be appreciated in such circuit where time-varying magnetic fields have been introduced, is that the 'electric potential difference' between any two points in the circuit is no longer uniquely defined (as we have seen in our earlier discussions of Sections A.6.3 to A.6.4). This is because since the line integral of the electric field around a closed contour in such a region is no-longer zero, the potential difference between any two (non-coincident) points on such contours (say points a and b or P and Q) is no longer independent of the path used to traverse from a to b or P to Q; i.e. the electric field is no longer conservative and the electric potential is no more a uniquely defined concept (since $\mathbf{E}$ cannot be expressed as '$-\nabla V$'). So we will now derive the equation for the behaviour of the current.

Hence consider a circuit as shown in Figure A.6.8 consisting of a single-turn loop, a battery and a resistor R both connected in series, with a switch S also in series with the other elements, and is closed at the instant $t = 0$.

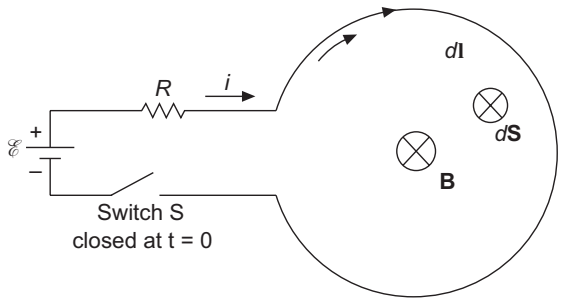


Figure A.6.8 Single-turn loop circuit.

For $t > 0$, a current would flow in the loop in the direction as shown in the figure (from the positive terminal of the battery to the negative). Then the Faraday's law is applied to this circuit, i.e. Eq. (A.6.28). The direction of traverse along the circuit is from positive plate of the battery to the negative plate (through the single-turn loop). Inside the battery, the direction of traverse (through the inside electric field of the battery) is against the field and hence its contribution to the line integral of **E** would be $-\mathscr{E}$. There is also an electric field inside the resistor R and so the contribution to $\mathbf{E} \cdot d\mathbf{l}$ would be $+iR$. If the loop wire has a small resistence $r << R$ then tis contribution to the line integral would be $+ir$ and this can be compounded with iR so that R now is redefined to include the external resistance as well as the resistance of the loop wire. Hence the L.H.S. of Eq. (A.6.28) is now

$$\oint \mathbf{E} \cdot d\mathbf{l} = -\mathscr{E} + iR \tag{A.6.29}$$

Next, the magnetic flux ϕ through the open surface bounded by the single turn loop is to be evaluated [for the R.H.S. of Eq. (A.6.28)] From the direction of the current round the loop, the flux ϕ is entering the plane of the paper normally which is also the assumed direction for $d\mathbf{S}$, and hence $\mathbf{B} \cdot d\mathbf{S}$ is +ve. It should be carefully noted that the magnetic field **B** in this region is due to the current in the loop and **is NOT** an externally imposed field. At any point in this region under consideration, **B** is proportional to the current i, as **B** can also be calculated by using Biot–Savart's law, i.e.

$$\mathbf{B} = i \oint \frac{\mu_0}{4\pi} \frac{d\mathbf{l} \times \mathbf{r}}{r^3} \tag{A. 6.30}$$

(The contribution to ϕ comes mostly from the single-turn loop, as the straight portion containing the battery and and the resistor R can be made as small as possible.) Hence it follows that for a given geometry of the loop, **B** is directly proportional to i. So ϕ can be expressed as

$$\phi = Li \tag{A.6.31}$$

where Li is a constant for a given geometry of the wire forming the loop, Li is called the self-inductance (or simply, the inductance) of the circuit. Hence Faraday's law gives

$$\oint \mathbf{E} \cdot d\mathbf{l} = -\mathscr{E} + iR = -\frac{d\phi}{dt} = -L\frac{di}{dt} \tag{A.6.32}$$

Thus the behaviour of the current is given by the equation

$$\frac{di}{dt} + \frac{R}{L} i = \frac{\mathscr{E}}{L} \tag{A.6.33a}$$

$$\therefore \qquad i(t) = \frac{\mathscr{E}}{R}\left\{1 - \exp\left(-\frac{R}{L}t\right)\right\} \qquad \text{(A.6.33b)}$$

where $\frac{L}{R} = \tau$ — characteristic time of the inductor = inductive time constant

This is the effect of non-zero inductance in the circuit.

A.6.5.1 Modified Form of Kirchhoff's Second Law (for Inductors)

Equation (A.6.32) for the current $i(t)$ can be rewritten as

$$\mathscr{E} - iR - L\frac{di}{dt} = 0 \qquad \text{(A.6.34)}$$

and this equation **looks like** that of Kirchhoff's second law, i.e. the sum of potential drops around a complete circuit is zero.

The rule 'implicit' in the derivation of the above equation is:

When an inductor is traversed in the direction of current-flow through it, the change in potential is $-L\frac{di}{dt}$. If the direction of travese is opposite to the direction of current flow, then the change in potential is $+L\frac{di}{dt}$.

While this approach produces the correct circuit equation for the current, the underlying physics of the phenomenon can be confusing. To say that there is a 'potential drop', across the inductor equal to $-L\frac{di}{dt}$, implies that the line integral of **E** through the inductor has a magnitude equal to $L\frac{di}{dt}$ which is based on the implicit assumption of an electric field existing in the inductor. But we have seen in the derivation of (A.6.32) that the line-integral of **E** through the one-loop inductor is zero (or nearly equal to zero). The question that needs to be looked into is that "even when there is a time-varying magnetic field here how does **E** field become zero"?

To answer this question, let there be a circular loop of wire, of radius a and total resistance R, which is positioned in a region where an externally imposed magnetic field, directed normally to the plane of the loop, exists and the field is increasing with time, as shown in Figure A.6.9.

In the present case, the externally imposed magnetic field (= B_{ext}) is assumed to be much greater than the magnetic field due to current in the wire and hence the internal field can be neglected. So only the effects of the external field need be considered.

The time-varying, externally imposed magnetic field gives rise to an induced electric field in the wire-loop.

The line integral of this induced electric field around the loop

$$= 2\pi a\ \mathbf{E}_{\text{induced}} = -\pi a^2\ \frac{d\mathbf{B}_{\text{ext}}}{dt} \leftarrow \text{ by Faraday's law}$$

or

$$\mathbf{E}_{\text{induced}} = -\frac{a}{2}\frac{d\mathbf{B}_{\text{ext}}}{dt}$$

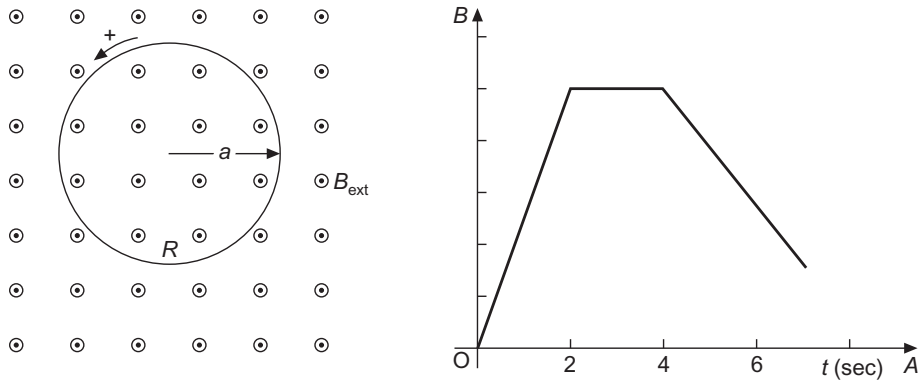


Figure A.6.9 A loop of wire positioned in an externally imposed time-varying magnetic field.

The direction of the induced electric field would be circumferential and it would be uniformly distributed along the contour of the circular loop (i.e. when only the contour is being considered or when the loop has open ends and has not been closed to allow any current to flow through it). But when the current is allowed to flow in the loop, the situation changes and becomes dependent on the resistance of the path along the circular wire-loop. If the resistance of the loop is uniformly distributed along the whole circumferential length (= $2\pi a$), then there is a uniform $\mathbf{E}_{\text{induced}}$ in the loop, this value being same at every point in the loop and circulating clockwise for $\mathbf{B}_{\text{ext}}$ increasing with time. This induced field causes a current to flow through the loop, the current density at every point by $\mathbf{J} = \mathbf{E}_{\text{induced}}/r$, r being the elemental resistance at every point of the loop. The total current in the uniform loop would be given by the total potential drop around the circular loop (i.e. P.D. = $2\pi a\mathbf{E}_{\text{induced}}$) divided by the total resistance of the loop (= R). Hence

$$I = \frac{2\pi a\, E_{\text{induced}}}{R}$$

and would circulate in the clockwire sense, i.e. same as that of $\mathbf{E}_{\text{induced}}$ for increasing $\mathbf{B}_{\text{ext}}$ [Figure A.610(a)]. Thus, when the resistance of the wire-loop is distributed uniformly along its circumference, then $\mathbf{E}_{\text{induced}}$ would be same at every point on the loop and would be circulating clockwise for $\mathbf{B}_{\text{ext}}$ increasing with time

The next question that needs looking into is, what happens when the resistance of the wire is not uniform along its length. The simplest situation of non-uniform resistance of the loop would be when the left-half of the loop has a resistance R_1 and the right-half has a resistance R_2 where $R_2 > R_1$ such that $R = R_1 + R_2$. Under this condition what would be the distribution of the $\mathbf{E}$ field in these two parts of the loop? Now the emf of the circuit is same as in the previous case as the total circuit resistance has not changed. But the electric field on the left-half of the loop (= $\mathbf{E}_1$) would be different from the right-half (= $\mathbf{E}_2$).

This must be so because the line-integral of the $\mathbf{E}$-field on the left side, over the left side = πaE_1 which would be equal to iR_1 (by Ohm's law). Similarly for the right side, $\pi aE_2 = iR_2$. Thus

$$\frac{E_1}{E_2} = \frac{R_1}{R_2}$$

and so $E_1 < E_2$ as $R_1 < R_2$. This should be correct as the loop wire would carry the same current, even though the resistances of the two halves are different. To maintain a unique value of the current through the whole loop, the E-field on two sides would get adjusted. Then what has

happened to the uniform electric field induced around the contour of the loop? It should be remembered that Faraday's law of electromagnetic induction still holds and the induced emf is a function of the time-changing magnetic field and does not depend on the resistance of different parts of the wire-loop positioned on the contour under consideration. What has been overlooked so far is that in the present situation there is another source of *E*-field apart from the field induced by the time-varying magnetic field. This second source of *E*-field is due to electric charges at the junctions separating wire segments of different resistances. Hence the resultant *E*-field in each of the two parts of the loop consists of (a) the induced *E*-field as a consequence of the Faraday's law, and (b) the charge field (= $\mathbf{E}_{\text{charge}}$) at the junctions directed from the +ve junction at the top [in Figure A.6.10(b)] to the –ve junction at the bottom of the loop.

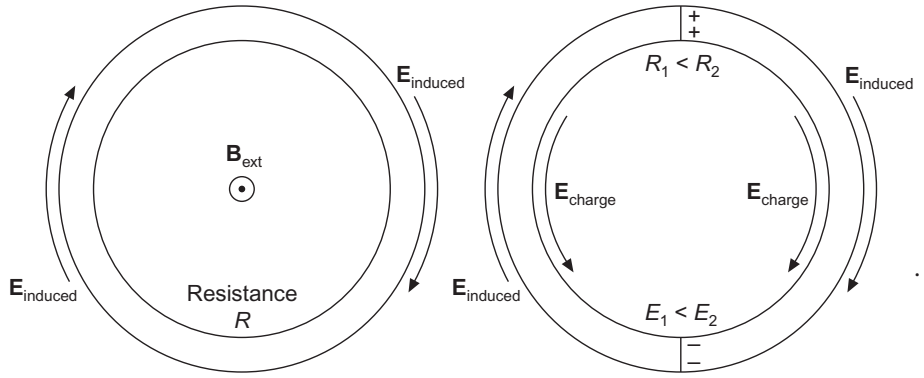


Figure A.6.10 Circular wire-loop in the externally imposed time-varying magnetic field.

The consequence of this field is to reduce the resultant electric field in the segment of lower resistance (= E_1) and increase the resultant *E*-field (= E_2) in the segment of higher resistance. This modification of the resultant *E*-field enables the same current to be maintained through the whole loop, as well as satisfy the requirement of the Faraday's law (i.e. the line integral of **E** over the whole loop be = $-\pi a^2 \frac{d\mathbf{B}_{\text{ext}}}{dt}$.)

Thus the non-uniform *E*-field in an inductor having non-uniform resistance along its length does not mean a breakdown of Faraday's law. According to Faraday's law, the induced *E* would be uniform over the 'whole circular contour' and the non-uniformity of the *E*-field is a consequence of the *E*-field produced by the junction charges.

If now a voltmeter is connected to the terminals of an inductor with very small resistance, what does the voltmeter indicate?

It needs a voltage drop of $L\frac{di}{dt}$. But this is not due to an electric field in the inductor. A current flows is the voltmeter because there is a large internal resistance in the voltmeter and the potential drop across that resistor $\left(\text{of } L\frac{di}{dt}\right)$, by Faraday's law applied to the voltmeter circuit, is what is read by the voltmeter.

REFERENCES

[1] Haus, H.A. and Melcher, J.R., 'Electric and Magnetic Fields', *Proc. IEEE*, Vol. **59**, No. 6, June 1971, pp. 887–894.

[2] Woodson H.H. and Melcher J.R., *Electromechanical Dynamics*, *Part I*, John Wiley and Sons, New York, 1968, pp. B9–13.

[3] Pramanik, A. et al., 'Circulating Current—Its Theory and Measurement', *BHEL Journal*, Vol. **13**, No. 3, 1992, pp. 26–32.

Additional Reference (Section A.6.4)

MIT Lecture Notes on Physics 8.02 by Prof. W. Lewin.

Appendix 7

Graphical Method of Analyzing Waveguide Problems (Smith Chart)

A.7.1 INTRODUCTION

Before the advent of computers, a number of graphical methods had been developed to solve or to assist in solving different types of waveguide problems. One of the most popular (if not the most widely used) methods has been the 'Smith Chart' which was developed by P.H. Smith in 1939, originally intended to analyze and design transmission line circuits. The initial intention of the inventor was to provide a handy graphical means for doing complex (variable) impedance calculations. This has now very wide applications in spite of the use of computers, and is now used as a presentation medium in computer-aided design (CAD) software for demonstrating the microwave circuit performances. Though there is a general impression that Smith chart is used for analyzing both lossy as well as loss-less transmission lines, it should be understood that it (Smith chart) can be used for solving all guided wave problems whose system terminal characteristics can be represented in terms of effective impedances. A further point to be noted is that even though in most of the elementary books on electromagnetism (including the present one), the transmission line problems are solved in terms of terminal impedances as circuit problems, it is possible to analyze the lines (either twin parallel lines, or eccentrically mounted parallel coaxial cables, or in general when the two parallel cylinders are **not** coaxial) in terms of their electromagnetic field distributions, by using a bi-cylindrical coordinate system (see Figure A.7.1). But since this coordinate system is not a part of undergraduate curriculum, the problem is treated as one of equivalent circuit analysis.

Since the Smith chart enables us to make a direct determination of the complex (voltage) reflection coefficient Γ (of the transmission line) by plotting in the complex plane, we shall generalize Γ for other waveguides such as rectangular waveguides. So before going into the principles of construction and usage of the chart, we shall extend the concept of the reflection coefficient to waveguides in general, and thus establish the validity of the chart to a wider range of problems.

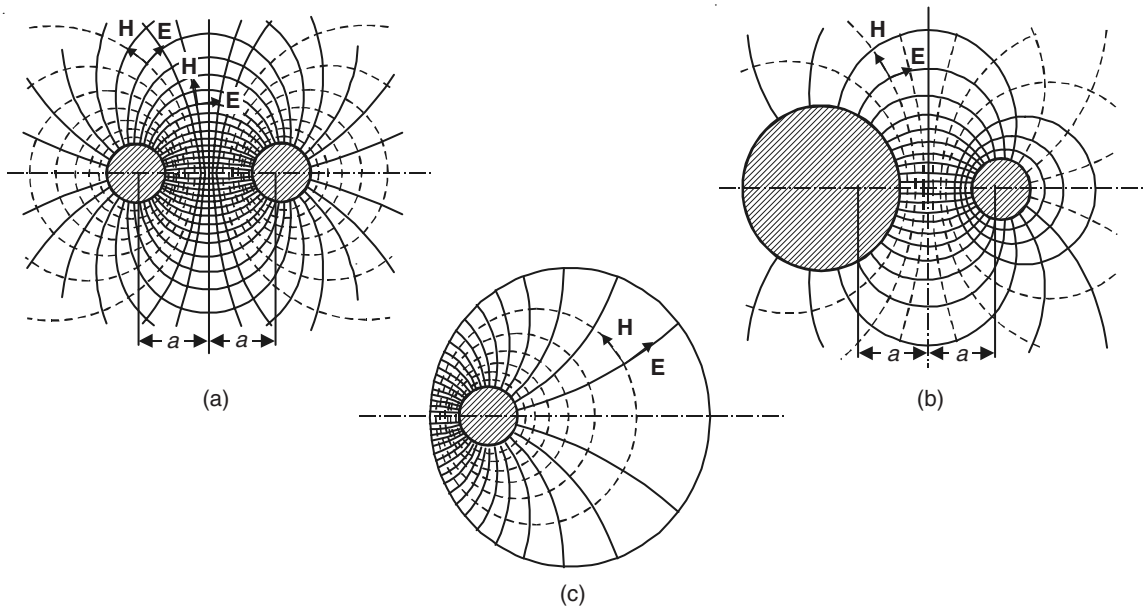


Figure A.7.1 TEM wave between parallel conductors. Bi-cylindrical coordinates are used for all three cases.

A.7.2 REFLECTION COEFFICIENT (VOLTAGE)

This has been already defined for lossless transmission lines on p. 598 as

$$\Gamma = \frac{Z_L - Z_c}{Z_L + Z_c} \tag{18.19}$$

where Z_L is the terminating load impedance and Z_c is the characteristic impedance of the line.

For the lossy line, we have denoted the reflection coefficient as Γ_R on p. 601, which is

$$\Gamma_R = \frac{Z_L - Z_c}{Z_L + Z_c} \qquad \text{[since the load impedance is } Z_L = Z_R \text{ of Eq. (18.31)]} \tag{18.31a}$$

It is also seen that in these equations, the numerator is proportional to the reflected voltage wave and the denominator to the incident wave, i.e.

$$\Gamma_R = \frac{V_2}{V_1}$$

Now we shall generalize this coefficient to include the behaviour of guided electric and magnetic waves (i.e. waveguides in TE and TM modes). For a TE wave, the coefficient Γ_e for the *E*-wave for (say) rectangular waveguides (as shown in Figure A.7.2) is

$$\Gamma_e = \frac{E_{x,\text{refl.}}}{E_{x,\text{inc.}}} \tag{A.7.1}$$

and for the corresponding magnetic wave:

$$\Gamma_m = \frac{H_{y,\text{refl.}}}{H_{y,\text{inc.}}} \tag{A.7.2}$$

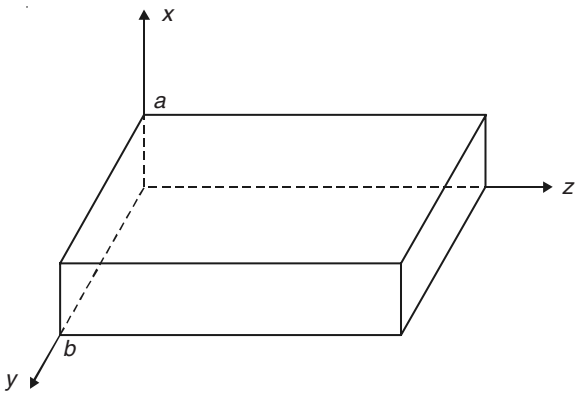


Figure A.7.2 Section of a rectangular waveguide.

For the simplicity of analysis, instead of considering a general *mn* mode of operation, we consider the simplest TE_{10} mode. In this mode, the **E** wave can be expressed as, starting with the incident wave ($E_{x,\text{ inc.}}$) as

$$E_{x,\text{ inc.}} = A \sin \frac{\pi y}{b} \exp(-j\beta z) \tag{A.7.3}$$

(*Note:* exp ($j\omega t$) is implicit in all these expressions.)

$$\therefore \quad -j\omega\mu H_y = \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} = -j\beta A \sin \frac{\pi y}{b} \exp(-j\beta z) \quad \text{(by Maxwell's equations)} \tag{A.7.4}$$

$$\therefore \quad H_{y,\text{ inc.}} = \frac{\beta}{\omega\mu} E_{x,\text{ inc.}} \tag{A.7.5}$$

Likewise for the reflected wave, we get

$$E_{x,\text{ refl.}} = B \sin \frac{\pi y}{b} \exp(+j\beta z) \tag{A.7.6}$$

and

$$H_{y,\text{ refl.}} = -\frac{\beta}{\omega\mu} E_{x,\text{ refl.}} \tag{A.7.7}$$

$$\therefore \quad \Gamma_m = \frac{H_{y,\text{ refl.}}}{H_{y,\text{ inc.}}} = -\frac{E_{x,\text{ refl.}}}{E_{x,\text{ inc.}}} = -\Gamma_e \tag{A.7.8}$$

So we shall use the notation, $\Gamma_e = \Gamma$ (A.7.9)

A.7.2.1 Variation of Reflection Coefficient along the Waveguide

We have the incident wave as

$$A \sin \frac{\pi y}{b} \exp(-j\beta z)$$

where z is the distance measured from the source (say). To express this in terms of the distance l from the load, we have

$$E_{x,\text{ inc.}} = A \sin \frac{\pi y}{b} \exp(+j\beta l) \tag{A.7.10}$$

Let the reflected wave set up at the load be of magnitude

$$\Gamma A \sin \frac{\pi y}{b}$$

Then at a point distant l from the load, this becomes

$$E_{x,\ \text{refl.}} = \Gamma A \sin \frac{\pi y}{b} \exp(-j\beta l) \tag{A.7.11}$$

(i.e. there is a relative phase lag)

$$E_{x,\ \text{total}} = A \sin \frac{\pi y}{b} \{\exp(+j\beta l) + \Gamma \exp(-j\beta l)\} \tag{A.7.12}$$

But the reflection coefficient Γ may be complex, i.e.

$$\Gamma = |\Gamma| \exp(j\phi) \tag{A.7.13}$$

where ϕ is the phase change of E on reflection, and a function of load impedance only.

Then Γ at the point $l = |\Gamma| \dfrac{\exp(j\phi) \exp(-j\beta l)}{\exp(+j\beta l)}$

$$\therefore \quad \Gamma = |\Gamma| \exp[j(\phi - 2\beta l)] \tag{A.7.14}$$

A.7.2.2 Standing Wave Pattern

We have

$$E_{x,\ \text{total}} = A \sin \frac{\pi y}{b} [\exp(+j\beta l) + \Gamma \exp(+j\beta l)]$$

$$= A \sin \frac{\pi y}{b} \exp(j\beta l)[1 + |\Gamma| \exp\{j(\phi - 2\beta l)\}] \quad \text{[from Eq. (A.7.13)]} \tag{A.7.15}$$

For graphical representation, since $|\Gamma| < 1$, we can consider a circle as shown in Figure A.7.3, in which $l = b$, $|\Gamma| = c$ and $E_{x,\text{total}} = a$.

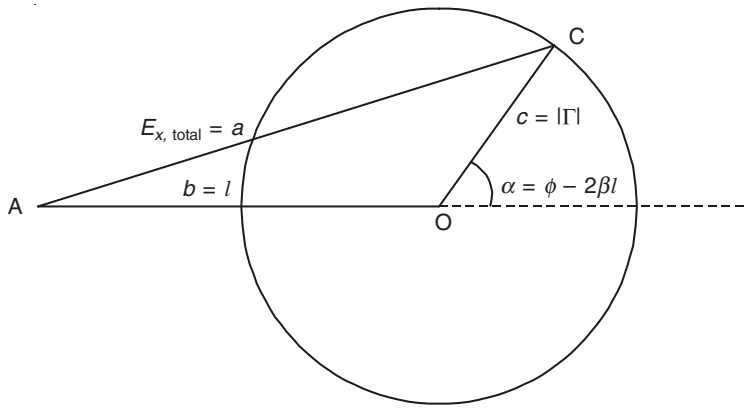


Figure A.7.3 Graphical representation of components of **E** wave.

From this ΔAOC, $\quad a^2 = (b + c \cos \alpha)^2 + c^2 \sin^2 \alpha$

$$= b^2 + c^2 + 2bc \cos \alpha, \qquad \text{where } \alpha = \phi - 2\beta l$$

$$\therefore \quad E^2_{x,\,\text{total}} = \left(A \sin \frac{\pi y}{b} \right)^2 \left\{ 1 + |\Gamma|^2 + 2|\Gamma| \cos (\phi - 2\beta l) \right\} \tag{A.7.16}$$

This has a maximum, when $\cos 2\left(\beta l - \frac{\phi}{2} \right) = 1$.

i.e. $$\beta l = \frac{\phi}{2} \tag{A.7.17}$$

$$\therefore \quad \text{The voltage maximum (for } E \text{ field)} = A \sin \frac{\pi y}{b} (1 + |\Gamma|) \tag{A7.18a}$$

and

$$\text{the } E\text{-field minimum} = A \sin \frac{\pi y}{b} (1 - |\Gamma|) \tag{A7.18b}$$

$$\therefore \quad \text{VSWR} = \frac{E_{\max}}{E_{\min}} = \frac{1 + |\Gamma|}{1 - |\Gamma|} = s \tag{A.7.19}$$

$$\therefore \quad |\Gamma| = \frac{s - 1}{s + 1} \tag{A.7.20}$$

A.7.2.3 Impedances

At any point in the wave, at a distance l from the load,

$$E_{x,\,\text{total}} = A \sin \frac{\pi y}{b} \exp (j\beta l)[1 + |\Gamma| \exp \{j(\phi - 2\beta l)\}] \tag{A.7.21}$$

$$H_{y,\,\text{total}} = \frac{A \sin \frac{\pi y}{b}}{Z_0} \exp (j\beta l)[1 - |\Gamma| \exp \{j(\phi - 2\beta l)\}] \tag{A.7.22}$$

(*Note:* The –ve sign in the above expression is due to magnetic Γ.)

and

$$Z_0 = \frac{\omega\mu}{\beta}$$

$$\therefore \quad Z_l = Z_0 \frac{1 + \Gamma_t \exp (-j2\beta l)}{1 - \Gamma_t \exp (-j2\beta l)} \tag{A.7.23}$$

and

$$Z_t = Z_0 \frac{1 + \Gamma_t}{1 - \Gamma_t} \qquad \text{and hence} \qquad \Gamma_t = \frac{Z_t - Z_0}{Z_t - Z_0} \tag{A.7.24}$$

(The subscript t above denotes the termination point at the load.)

$\therefore$ From (A.7.23) and (A.7.24), impedance at any point ($= Z_l$) is

$$Z_l = Z_0 \frac{(Z_t + Z_0) \exp (j\beta l) + (Z_t - Z_0) \exp (-j\beta l)}{(Z_t + Z_0) \exp (j\beta l) - (Z_t - Z_0) \exp (-j\beta l)}$$

$$= Z_0 \frac{Z_t + jZ_0 \tan \beta l}{Z_0 + jZ_t \tan \beta l} \tag{A.7.25}$$

$$\therefore \qquad \text{Normalized impedance, } Z_l' = \frac{Z_1}{Z_0} \tag{A.7.26a}$$

and

$$\text{Normalized admittance, } Y_l' = \frac{(1/Z_l)}{(1/Z_0)} \tag{A.7.26b}$$

The vectorial representations of **E** and **H** are shown in Figure A.7.4.

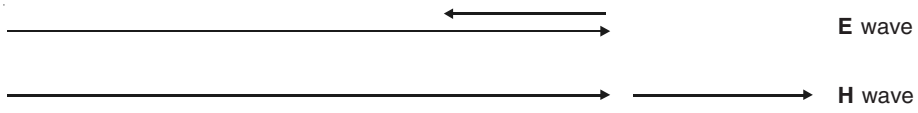


Figure A.7.4 Vectorial representation of **E** and **H** waves.

A.7.3 SMITH CHART

As mentioned earlier, the Smith chart helps us to make a direct determination of the complex reflection coefficient for a given load impedance Z_L and the characteristic impedance Z_0. Hence we start with

$$\Gamma_l \text{ [or } \Gamma(l)] = \Gamma_t \exp(-j2\beta l) \tag{A.7.27}$$
(= reflection coeff. at the point l)

Any reflection coefficient Γ can be represented by a point on the graph (of the complex plane).

$$\therefore \quad Z_l' = \frac{Z_l}{Z_0} = \frac{1 + |\Gamma| \exp\left[j(\phi - 2\beta l)\right]}{1 - |\Gamma| \exp\left[j(\phi - 2\beta l)\right]} = \frac{1 + \Gamma(l)}{1 - \Gamma(l)} = r + jx \tag{A.7.28}$$

(*Note:* Here r is the real part and not to be confused with the r-axis of polar coordinate system.)

Let

$$\Gamma(l) = u + jv \tag{A.7.29}$$

$$\therefore \quad r + jx = \frac{1 + u + jv}{1 - u - jv} = \frac{(1 - u^2 - v^2) + j2v}{(1 - u)^2 + v^2}$$

$$\therefore \quad r = \frac{1 - u^2 - v^2}{(1 - u)^2 + v^2}, \qquad x = \frac{2v}{(1 - u)^2 + v^2} \text{ (= reactance component)} \tag{A.7.30}$$

Considering the real part,

$$[(1 - u)^2 + v^2]r = 1 - u^2 - v^2$$

or

$$u^2 - \frac{2ur}{r+1} + \frac{r-1}{r+1} + v^2 = 0$$

$$\therefore \qquad \left(u - \frac{r}{r+1}\right)^2 + v^2 = \left(\frac{r}{r+1}\right)^2 - \frac{r-1}{r+1} = \left(\frac{1}{r+1}\right)^2 \tag{A.7.31}$$

The above equation represents a family of coaxial circles with centres at $\left(\frac{r}{r+1}, 0\right)$, and radii $\frac{1}{r+1}$ (Figure A.7.5).

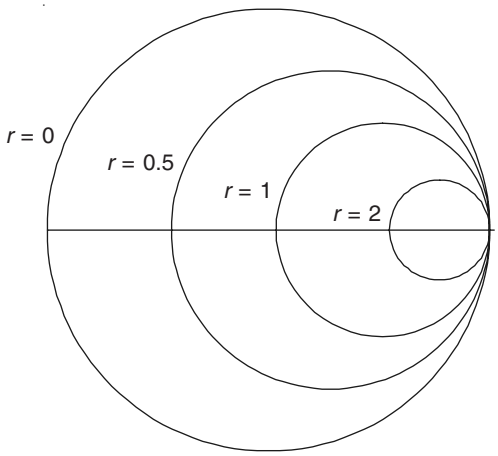


Figure A.7.5 Representation of real part of $\Gamma(l)$ in Smith chart.

Next considering the coefficients of j from Eq. (A.7.30)

$$(u - 1)^2 + v^2 = \frac{2v}{x}$$

or

$$(u - 1)^2 + \left(v - \frac{1}{x}\right)^2 = \left(\frac{1}{x}\right)^2 \tag{A.7.32}$$

These take the form of circles, for different values of x, of centres $(1/x)$, x being +ve or –ve, and radii $1/|x|$ (Figure A.7.6).

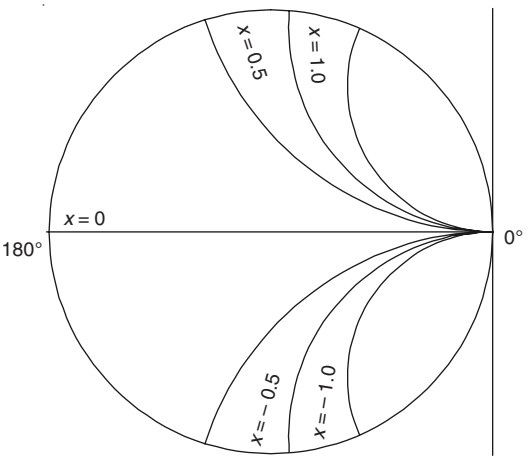


Figure A.7.6 Representation of coefficient of j in Smith chart.

These two families of circles, superimposed form the Smith chart. The charts can be of both impedances as well as admittances, as shown in Figures A.7.7 and A.7.8.

Some examples with solutions are given below for practice:

1. A normalized impedance $Z'_l = 0.5 + j0.5$
corresponds to a reflection coefficient 0.45 exp ($j115$).

2. Calculate the impedance at a distance of 30° from the previous point, closer to the generator = $1.4 + j1.1$.

Check from the formula, $Z'_l = \dfrac{Z'(l) + j \tan \beta (l_2 - l_1)}{1 + Z'(l) \tan \beta (l_2 - l_1)}$

$$\pi = \frac{\lambda}{4} \qquad \text{and} \qquad 30° = \frac{\lambda}{24}$$

3. VSWR = 0.25, the first voltage minimum is 0.3 cm from the load, λ_g = 3 cm. Find the load.

Procedure: Enter the Smith chart at the voltage minimum at VSWR = 0.25, ϕ corresponds to $\dfrac{\lambda}{10} = \dfrac{0.3 \text{ cm}}{3 \text{ cm}}$, $\lambda = 4\pi°$

$\therefore \quad \dfrac{\lambda}{10} \rightarrow 72°$ on Smith chart towards the load.

This gives $Z'_l = \dfrac{Z_l}{Z_0} = 0.37 - j0.66$

Admittance corresponding to this impedance is given by joining the point to the origin, and continuing the line an equal distance beyond the origin.

In this case, the impedance = $0.7 + j1.1 = Y'_l = \dfrac{Y_l}{Y_0}$.

The Complete Smith Chart

Black Magic Design

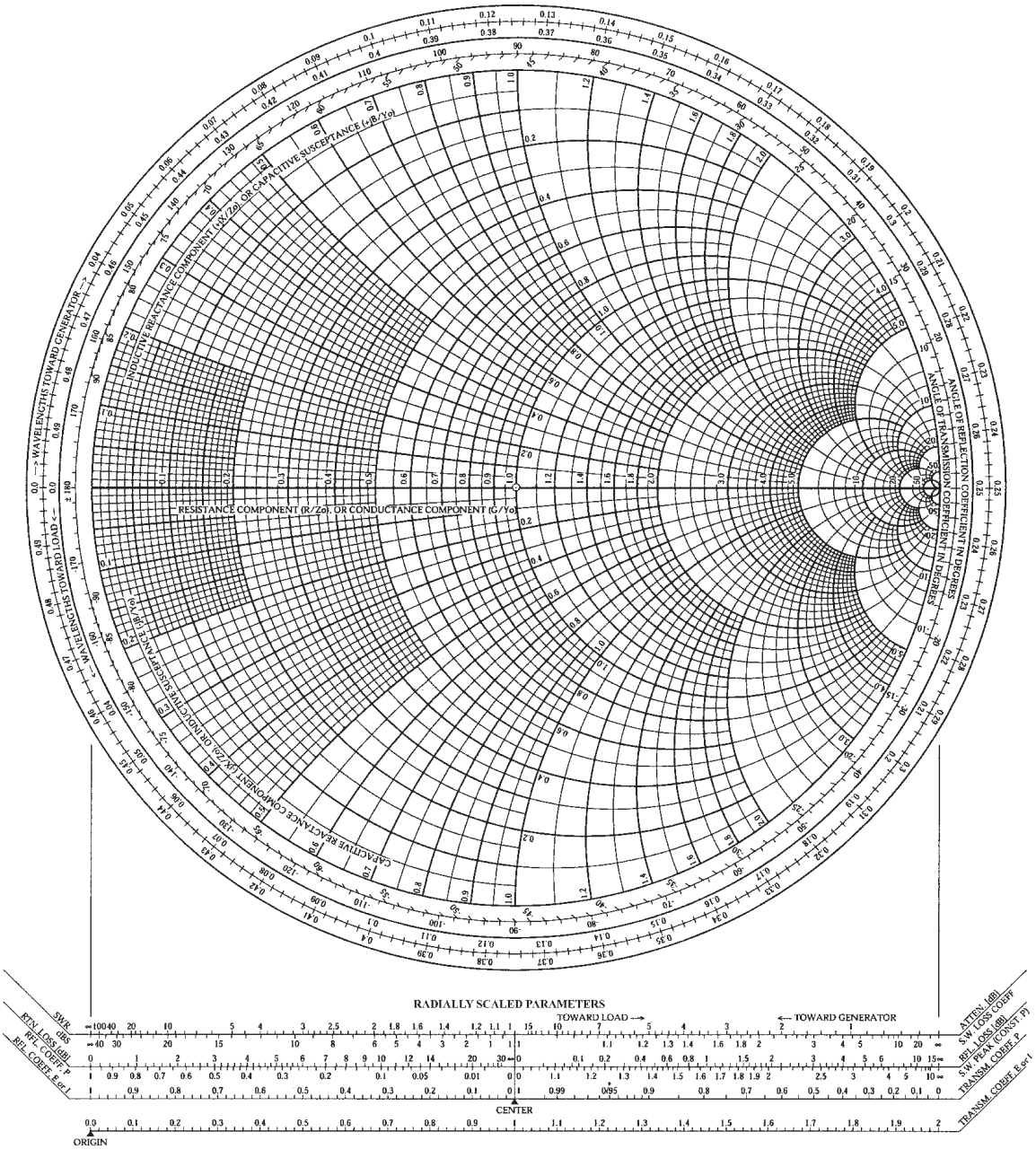


Figure A.7.7 Impedance or admittance coordinates.

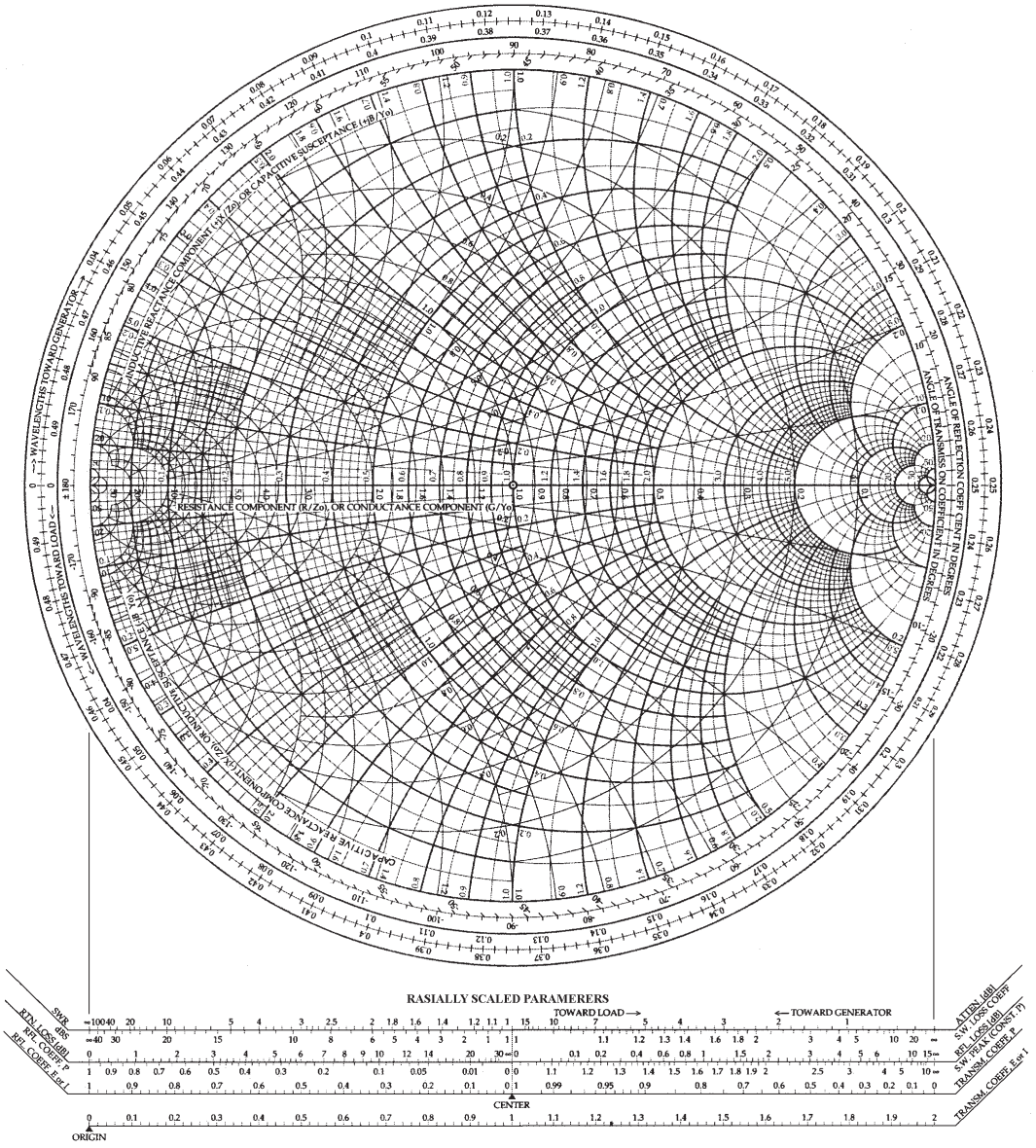


Figure A.7.8 Normalized impedance and admittance coordinates.

Appendix 8

A Proof of the Fact that There are only Two Independent Boundary Conditions for Electromagnetic Field Problems

(A)

In electrostatics, on an interface of discontinuity between two media of different permittivities, the conditions are

$$\left.\begin{aligned} \hat{\mathbf{n}} \cdot (\mathbf{D_1} - \mathbf{D}) &= \rho_S \\ \text{and} \\ \hat{\mathbf{n}} \times (\mathbf{E}_1 - \mathbf{E}_2) &= 0 \end{aligned}\right\} \tag{A.8.1}$$

where $\hat{\mathbf{n}}$ is the unit vector normal to the interface plane separating the two media, and $\mathbf{D}_1$, $\mathbf{D}_2$ and $\mathbf{E}_1$, $\mathbf{E}_2$ are the relevant electrostatic field vectors (i.e. the electric flux density and the electric field intensity respectively), and ρ_S is the charge density on the interface plane.

Similarly, in magnetostatics, on an interface of discontinuity between two media of different permeabilities, the conditions are

$$\left.\begin{aligned} \hat{\mathbf{n}} \times \left(\mathbf{H}_1 - \mathbf{H}_2\right) &= \mathbf{J}_S \\ \text{and} \\ \hat{\mathbf{n}} \cdot \left(\mathbf{B}_1 - \mathbf{B}_2\right) &= 0 \end{aligned}\right\} \tag{A.8.2}$$

where, as before $\hat{\mathbf{n}}$ is the normal unit vector, and $\mathbf{H}_1$, $\mathbf{H}_2$ and $\mathbf{B}_1$, $\mathbf{B}_2$ are the relevant magnetostatic field vectors (i.e. the magnetic intensity vector and the magnetic flux density vector respectively), and $\mathbf{J}_S$ is the surface current density on the interface plane.

So we take a surface S and consider a small area ΔA on it (Figure A.8.1),

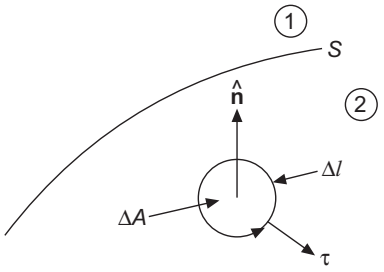


Figure A.8.1 The interface surface S and a small elemental area ΔA on S.

where

Δl = the contour enclosing ΔA

$\hat{\mathbf{n}}$ = the outward (normal to S) unit vector

τ = the outward (normal to Δl) unit vector, and tangential to S.

We define:

$$\left.\begin{aligned} &\Delta\mathbf{H} = \mathbf{H}_1 - \mathbf{H}_2, \qquad \Delta\mathbf{D} = \mathbf{D}_1 - \mathbf{D}_2; \\ \text{and} \\ &\Delta\mathbf{J}_v = \mathbf{J}_{v_1} - \mathbf{J}_{v_2} \end{aligned}\right\} \tag{A.8.3}$$

where $\mathbf{J}_{v_1}$ and $\mathbf{J}_{v_2}$ are the volume current densities on two sides of the interface surface S under consideration.

Now, we have $$\iint_S (\text{curl } \mathbf{A}) \cdot \mathbf{n}\, da = \int_C \mathbf{A} \cdot d\mathbf{S},$$

where S is the surface enclosed by the contour C.

In the present problem,

$$\begin{aligned} \iint_{\Delta A} \{\nabla \times \Delta\mathbf{H}\} \cdot \mathbf{n}\, dS &= \int_{\Delta l} \{\Delta\mathbf{H} \cdot \hat{\mathbf{l}}\}\, dl \qquad (\hat{\mathbf{l}} \text{ is the unit vector in the direction of } \Delta l) \\ &= \iint_{\Delta A} (j\omega\Delta\mathbf{D} + \Delta\mathbf{J}_v) \cdot \hat{\mathbf{n}}\, dS \quad \text{by Maxwell's equations} \\ &= \int_{\Delta l} (\Delta\mathbf{H}_t) \cdot \hat{\mathbf{l}}\, dl \end{aligned} \tag{A.8.4}$$

where $\Delta\mathbf{H}_t$ is the tangential component of $\Delta\mathbf{H}$ and $j\omega = \dfrac{\partial}{\partial t}$

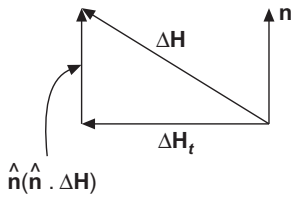


Figure A.8.2 Directions of the unit vectors.

Now, $\Delta\mathbf{H}_t = -\hat{\mathbf{n}} \times (\hat{\mathbf{n}} \times \Delta\mathbf{H}) = \Delta\mathbf{H} - \hat{\mathbf{n}}(\hat{\mathbf{n}} \cdot \Delta\mathbf{H})$

$= -\hat{\mathbf{n}} \times \mathbf{J}_S$, $\mathbf{J}_S$ being the surface current density on the interface surface S.

From Eq. (A.8.4),

$$\iint_{\Delta A} j\omega\, \Delta\mathbf{D} \cdot \hat{\mathbf{n}}\, dS = -\iint_{\Delta A} \Delta\mathbf{J}_v \,.\, \hat{\mathbf{n}}\, dS - \oint_{\Delta l} (\hat{\mathbf{n}} \times \mathbf{J}_S) \cdot \hat{\mathbf{l}}\, dl \tag{A.8.5}$$

We take the last term to change it into an area integral,

$$(\hat{\mathbf{n}} \times \mathbf{J}_S) \cdot \hat{\mathbf{l}} = \hat{\mathbf{l}} \cdot (\hat{\mathbf{n}} \times \mathbf{J}_S) = (\hat{\mathbf{l}} \times \hat{\mathbf{n}}) \cdot \mathbf{J}_S = \hat{\boldsymbol{\tau}} \,.\, \mathbf{J}_S$$

$\int_{\Delta l} (\hat{\tau} \,.\, \mathbf{J}_S)\, dl$ is the measure of the divergence or outward flow of charge from the region ΔA.

From the continuity of current and charge,

$$\int_{\Delta l} (\hat{\boldsymbol{\tau}} \,.\, \mathbf{J}_S)\, dl = -\frac{\partial}{\partial t} \iint_{\Delta A} q'_S dS = -j\omega \iint q'_S dS = \int_{\Delta l} (\hat{\mathbf{n}} \times \mathbf{J}_S) \cdot \hat{\mathbf{l}}\, dl \tag{A.8.6}$$

where q'_S is the charge generated by the surface divergence of the current flow.

$$\iint_{\Delta A} (\Delta\mathbf{D} \cdot \hat{\mathbf{n}})\, dS = -\iint_{\Delta A} \frac{\Delta\mathbf{J}_v \cdot \hat{\mathbf{n}}}{j\omega}\, dS = \iint_{\Delta A} q_S\, dS \tag{A.8.7}$$

where $-\dfrac{\Delta\mathbf{J}_v \cdot \hat{\mathbf{n}}}{j\omega} = q_S$, the surface charge density due to the volume divergence of the current flow.

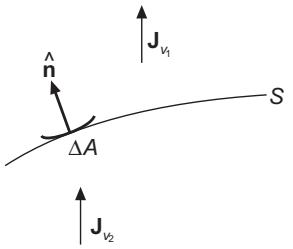


Figure A.8.3 Current density vectors.

$$\rho_S = q'_S + q_S \tag{A.8.8}$$

$$\therefore \quad \iint_{\Delta A} \Delta\mathbf{D} \cdot \hat{\mathbf{n}}\, dS = \iint \rho_S\, dS \tag{A.8.9}$$

This is true for any ΔA; hence

$$\Delta\mathbf{D} \cdot \hat{\mathbf{n}} = \rho_S$$

or $$\hat{\mathbf{n}} \cdot (\mathbf{D}_1 - \mathbf{D}_2) = \rho_S$$

as a consequence of $$\hat{\mathbf{n}} \times (\mathbf{H}_1 - \mathbf{H}_2) = \mathbf{J}_S \tag{A.8.10}$$

Similarly it can be shown that

$$\hat{\mathbf{n}} \cdot (\mathbf{B}_1 - \mathbf{B}_2) = 0 \text{ is a consequence of } \hat{\mathbf{n}} \times (\mathbf{E}_1 - \mathbf{E}_2) = 0 \tag{A.8.11}$$

$\therefore$ The essential boundary conditions are:

$$\hat{\mathbf{n}} \times (\mathbf{E}_1 - \mathbf{E}_2) = 0 \quad \text{and} \quad \hat{\mathbf{n}} \times (\mathbf{H}_1 - \mathbf{H}_2) = \mathbf{J}_S \tag{A.8.12}$$

and these are generally the best to use.

For metals only, $$\hat{\mathbf{n}} \times \mathbf{E}_1 = 0 \quad \text{and} \quad \hat{\mathbf{n}} \times \mathbf{H}_1 = \mathbf{J}_S \tag{A.8.13}$$

For dielectrics, $$\hat{\mathbf{n}} \times (\mathbf{E}_1 - \mathbf{E}_2) = 0 \quad \text{and} \quad \hat{\mathbf{n}} \times (\mathbf{H}_1 - \mathbf{H}_2) = \mathbf{J}_S \tag{A.8.14}$$

Note: This proof is due to Dr. J.R. Wait, then of N.B.S. and was given to me by Dr. T.S.M. McLean, Reader in Electromagnetism, Department of Electronic & Electrical Engineering, University of Birmingham, who had got it from Dr. Wait while attending one of the summer schools at Washington.

(B) A COMPACT FORM OF THE PROOF OF THE PREVIOUS THEOREM

Starting from the Continuity Equation,
(and using the differential notation, while remembering that all these equations have been derived from their integral forms and hence do not require the continuity implicit in the differential form of these equations)

$$\nabla \cdot \mathbf{J} = \frac{\partial \rho}{\partial t} = -j\omega\rho$$

$$\nabla \cdot (\mathbf{J}_n - \mathbf{J}_t)\big]_{\text{surface}} = -j\omega\rho_S$$

or

$$\underbrace{\nabla \cdot \mathbf{J}_n\big]_{\text{surface}}}_{\downarrow} + \nabla \cdot \mathbf{J}_t\big]_{\text{surface}} = -j\omega\rho_S$$

$$J_{n_1} - J_{n_2} + \nabla \cdot \mathbf{J}_S = -j\omega\rho_S$$

From the Maxwell's equations,

$$\nabla \times (\mathbf{H}_1 - \mathbf{H}_2) = j\omega(\mathbf{D}_1 - \mathbf{D}_2) + (\mathbf{J}_1 - \mathbf{J}_2)$$

or

$$\mathbf{n} \cdot \{\nabla \times (\mathbf{H}_1 - \mathbf{H}_2)\} = j\omega(\mathbf{D}_{1n} - \mathbf{D}_{2n}) + (\mathbf{J}_{1n} - \mathbf{J}_{2n})$$

Using a vector identity for the L.H.S., we get:

$$\nabla \cdot \{(\mathbf{H}_1 - \mathbf{H}_2) \times \mathbf{n}\} + (\mathbf{H}_1 - \mathbf{H}_2) \cdot (\nabla \times \mathbf{n}) = j\omega(\mathbf{D}_{1n} - \mathbf{D}_{2n}) + (\mathbf{J}_{1n} - \mathbf{J}_{2n})$$

Now, $$\nabla \times \mathbf{n} = 0$$

{*Note:* The given boundary condition is $\mathbf{n} \times (\mathbf{H}_1 - \mathbf{H}_2) = \mathbf{J}_S$}

From the above equation, it reduces to

$$\nabla\cdot(-\mathbf{J}_S) = j\omega(D_{1n} - D_{2n}) + (J_{1n} - J_{2n})$$

or

$$-[\nabla \times \mathbf{J}_S + (J_{1n} - J_{2n})] = j\omega(D_{1n} - D_{2n})$$

or

$$j\omega r_S = j\omega(D_{1n} - D_{2n})$$

If $\omega \neq 0$,

$$\rho_S = D_{1n} - D_{2n}$$ Q.E.D.

Note: This compact proof was derived by Prof. Eugene Kopp of the University of California who was then a Visiting Fellow in the Department of Electrical & Electronics Engineering, the University of Leeds from the rigourous proof derived by Dr. J.R. Wait. Prof. Kopp had borrowed the Wait's proof from me and in return gave me this proof a few days later.

Appendix 9

Bessel Functions and Legendre Functions

A.9.1 INTRODUCTION

Bessel functions of different kinds are used in cylindrical geometry problems of electrostatic fields (Chapter 4), magnetic fields as well as of electromagnetic waves (e.g. cylindrical waveguides in Chapter 18). Legendre functions are used in problems of spherical geometry over the whole range of electromagnetism [i.e. electrostatic fields in spherical capacitors to those of antennae (Chapter 19)]. It should be appreciated that the use of these functions for these two types of geometries is not on any arbitrary basis but because these two functions are the solutions of the differential equations obtained by the method of separation of variables from the Laplace's equation (or more generally from the Laplacian operator of the more general type of Helmholtz equation) in these two coordinate systems. The derivation of two ODE's (i.e. Bessel's equations and Legendre equation) has been given in Chapter 4 (Sections 4.2.5 and 4.2.7). So we will briefly summarize the expressions for the series solutions, their various properties and their recurrence relations used for simplifying the expressions algebraically. For both the equations, the series solutions are represented by functions having an order and an argument.

A.9.2 BESSEL FUNCTIONS

Bessel functions are solutions of the ordinary differential equation of the second order (known as Bessel's equation),

$$\frac{d^2R}{dr^2} + \frac{1}{r}\frac{dR}{dr} + \left(k_z^2 - \frac{k_\phi^2}{r^2}\right) R = 0$$

where $$R = R(r)$$

Its solutions are Bessel's functions of the first kind $J_{k_\phi}(k_z r)$ and of the second kind $Y_{k_\phi}(k_z r)$, both of order k_ϕ and argument $k_z r$

$$R(r) = A\, J_{k_\phi}(k_z r) + B\, Y_{k_\phi}(k_z r)$$

where $$J_{k_\phi}(k_z r) = \sum_{r=0}^{\infty} (-1)^r \frac{\left(k_z r/2\right)^{k_\phi + 2r}}{\Gamma(r+1)\,\Gamma(k_\phi + r + 1)}$$

where Γ is a gamma function defined as

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt = (x-1)\,\Gamma(x-1) = (x-1)!$$

and

$$Y_{k_\phi}(k_z r) = \frac{\cos(k_\phi \pi)\, Y_{k_\phi}(k_z r) - J_{-k_\phi}(k_z r)}{\sin(k_\phi \pi)}$$

For k_ϕ integer $\quad J_{-k_\phi}(k_z r) = (-1)^{k_\phi} J_{k_\phi}(k_z r)$

where the general solution of the Laplace's equation is:

$$V(r, \phi, z) = \sum_{k_\phi} \sum_{k_z} \left\{ A J_{k_\phi}(k_z r) + B Y_{k_\phi}(k_z r) \right\} \left\{ C \cos(k_\phi \phi) + D \sin(k_\phi \phi) \right\} \left\{ E \cosh(k_z z) + F \sinh(k_z z) \right\}$$

and when $k_z = 0$

$$V_0 = \sum_{k_\phi} \left(A r^{k_\phi} + B r^{-k_\phi} \right) \left\{ C \cos(k_\phi \phi) + D \sin(k_\phi \phi) \right\} \left\{ E_0 z + F_0 \right\}$$

and when both $k_z = 0, \quad k_\phi = 0$

$$V_{oo} = (A_{oo} \ln r + B_{oo})(C_o \phi + D_o)(E_o z + F_o)$$

Notes: 1. $Y_{k_\phi}(k_z r)$ is also known as the "Neumann function".
2. Bessel's functions in general terms are also called cylinder functions.

When the constant k_z^2 has a negative sign, the differential equation becomes:

$$\frac{d^2 R}{dr^2} + \frac{1}{r}\frac{dR}{dr} + \left(-k_z^2 - \frac{k_\phi^2}{r^2} \right) R = 0$$

This equation is called the modified Bessels equation and its solutions are the modified Bessel function of the first kind $I_{k_\phi}(k_z r)$ and of the second kind $K_{k_\phi}(k_z r)$, both of order k_ϕ and argument $(k_z r)$, which are

$$I_{k_\phi}(k_z r) = \sum_{r=0}^{\infty} \frac{(k_z r/2)^{k_\phi + 2r}}{\Gamma(r+1)\,\Gamma(k_\phi + r + 1)} = j^{-k_\phi} J_{k_\phi}(j k_z r)$$

$$K_{k_\phi}(k_z r) = \frac{\pi}{2} \left\{ \frac{I_{-k_\phi}(k_z r) - I_{k_\phi}(k_z r)}{\sin(k_\phi \pi)} \right\}$$

For k_ϕ integer, $\quad I_{-k_\phi}(k_z r) = j^{k_\phi} J_{-k_\phi}(j k_z r) = I_{k_\phi}(k_z r)$

So we will write these two equations in the more general terms in y and x, i.e.

$$\frac{d^2 y}{dx^2} + \frac{1}{x}\frac{dy}{dx} + \left(1 - \frac{n^2}{x^2} \right) y = 0$$

whose solutions are $\quad y = A J_n(x) + B Y_n(x)$

and

$$\frac{d^2y}{dx^2} + \frac{1}{x}\frac{dy}{dx} - \left(1 + \frac{n^2}{x^2}\right) y = 0$$

whose solutions are $y = AI_n(x) + BK_n(x)$

(These functions are of the order n and argument x.)

Next, we will first consider the properties of (ordinary) Bessel functions.

A.9.3 ORDINARY BESSEL FUNCTIONS

A.9.3.1 Asymptotic Forms

As $n \to \infty$ through real +ve values, other variables remaining constant

$$J_n(x) \sim \frac{1}{\sqrt{2\pi n}} \left(\frac{ex}{2n}\right)^n$$

$$Y_n(x) \sim -\sqrt{\frac{2}{\pi n}} \left(\frac{ex}{2n}\right)^{-n}$$

Limiting forms for small arguments; when n is fixed and $x \to 0$

$$J_n(x) \sim \frac{(x/2)^n}{\Gamma(n+1)} \qquad (n \neq -1, -2, -3, \ldots)$$

$$Y_n(x) \sim \frac{1}{\pi}\Gamma(n)\,(x/2)^{-n} \qquad \text{Real } n > 0$$

In particular,

$$J_o(x) \sim 1$$

$$Y_o(x) \sim \frac{2}{\pi}\left(\ln x + C - \ln 2\right)$$

where C is constant.

For $x >> 1$, the asymptotic expressions are

$$J_n(x) \sim \sqrt{\frac{2}{\pi x}} \cos\left(x - \frac{n\pi}{2} - \frac{\pi}{4}\right)$$

$$Y_n(x) \sim \sqrt{\frac{2}{\pi x}} \sin\left(x - \frac{n\pi}{2} - \frac{\pi}{4}\right)$$

A.9.3.2 Differentiation and Integration of Bessel Functions

From the series definition of these functions,

$$J_o'(x) = -J_1(x)$$

and

$$Y_o'(x) = -Y_1(x)$$

If we use C_n to denote J_n or Y_n or $H_n^{(1)}$ or $H_n^{(2)}$ {i.e. Hankel functions of the two kinds (to be defined later)} or any linear combination of these functions,

$$C_0'(x) = C_1(x)$$

$$C_1'(x) = C_0(x) - \frac{1}{x} C_1(x)$$

More generally

$$\left(\frac{1}{x}\frac{d}{dx}\right)^k \left\{x^n C_n(x)\right\} = x^{n-k} C_{n-k}(x)$$

$$\left(\frac{1}{x}\frac{d}{dx}\right)^k \left\{x^{-n} C_n(x)\right\} = (-1)^k x^{-n-k} C_{n+k}(x)$$

The integrals are:

$$\int_0^x C_1(t)\, dt = 1 - C_0(x)$$

$$\int_0^x C_{2n}(t)\, dt = \int_0^x C_0(t)\, dt - 2\sum_{k=0}^{2n-1} J_{2k+1}(x)$$

$$\int_0^x C_{2n+1}(t)\, dt = 1 - C_0(x) - 2\sum_{k=1}^{n} J_{2k}(x)$$

$$\int_0^x C_{n+1}(t)\, dt = \int_0^x C_{n-1}(t)\, dt - 2C_n(x), \quad n > 0$$

A.9.3.3 Wronskians

$$W\left\{J_n(x), J_{-n}(x)\right\} = J_{n+1}(x)\, J_{-n}(x) + J_n(x)\, J_{-(n+1)}(x)$$

$$= -\frac{2\sin(n\pi)}{\pi x}$$

$$W\left\{J_n(x), Y_n(x)\right\} = J_{n+1}(x)\, Y_n(x) - J_n(x)\, Y_{(n+1)}(x)$$

$$= \frac{2}{\pi x}$$

A.9.3.4 Recurrence Relations

$$C_{n-1}(x) + C_{n+1}(x) = \frac{2n}{x} C_n(x)$$

$$C_{n-1}(x) - C_{n+1}(x) = 2C_n'(x)$$

$$C_n'(x) = C_{n-1}(x) - \frac{n}{x} C_n(x)$$

$$= - C_{n+1}(x) + \frac{n}{x} C_n(x)$$

A.9.4 MODIFIED BESSEL FUNCTIONS

A.9.4.1 Asymptotic Forms

For n fixed and $x \to 0$

$$I_n(x) \sim \frac{(x/2)^n}{\Gamma(n+1)}, \qquad n \neq -1, -2, \ldots$$

$$K_0(x) \sim - \ln x$$

$$K_n(x) \sim \frac{1}{2} \Gamma(n) \left(x/2\right)^{-n}$$

When $x >> 1$, $I_n(x) \to \infty$ and $K_n(x) \to 0$.

A.9.4.2 Wronskians

$$W\left\{I_n(x), I_{-n}(x)\right\} = I_n(x)\, I_{-(n+1)}(x) - I_{n+1}(x)\, I_{-n}(x)$$

$$= - \frac{2 \sin(n\pi)}{\pi x}$$

$$W\left\{K_n(x), I_n(x)\right\} = I_n(x)\, K_{n+1}(x) + I_{n+1}(x)\, K_n(x)$$

$$= \frac{1}{x}$$

A.9.4.3 Recurrence Relations

$$I_{n-1}(x) - I_{n+1}(x) = \frac{2n}{x} I_n(x)$$

$$I_n'(x) = I_{n-1}(x) - \frac{n}{x} I_n(x) = I_{n+1}(x) + \frac{n}{x} I_n(x)$$

$$I_{n-1}(x) + I_{n+1}(x) = 2 I_n'(x)$$

$$K_n'(x) = \frac{n}{x} K_n(x) - K_{n+1}(x) = -K_{n-1}(x) - \frac{n}{x} K_n(x)$$

$$I_0'(x) = I_1(x), \quad \text{and} \quad K_0'(x) = -K_1(x)$$

$$\left(\frac{1}{x}\frac{d}{dx}\right)^k \left\{x^n\ C_n(x)\right\} = x^{n-k}\ C_{n-k}(x)$$

$$\left(\frac{1}{x}\frac{d}{dx}\right)^k \left\{x^{-n}\ C_n(x)\right\} = x^{-n-k}\ C_{n+k}(x)$$

A.9.5 BESSEL FUNCTIONS OF THE THIRD KIND

These solutions are also called Hankel functions and are solutions of the ordinary Bessel's equation. They are defined as

$$H_n^{(1)}(x) = J_n(x) + j\,Y_n(x)$$
$$= j\left\{\text{cosec}\,(n\pi)\right\}\left\{\exp(-jn\pi)\ J_n(x) - J_{-n}(x)\right\}$$
$$H_n^{(2)}(x) = J_n(x) - j\,Y_n(x)$$
$$= j\left\{\text{cosec}\,(n\pi)\right\}\left\{J_{-n}(x) - \exp(-jn\pi)\ J_n(x)\right\}$$

These are Hankel functions of the first and second kind respectively.

$$H_{-n}^{(1)}(x) = \left\{\exp(jn\pi)\right\} H_n^{(1)}(x) \quad \text{and} \quad H_{-n}^{(2)}(x) = \left\{\exp(-jn\pi)\right\} H_n^{(2)}(x)$$

The recurrence relations of ordinary Bessel functions (i.e. Section A.9.3.4) apply to Hankel functions.

These functions are useful in solving eddy current problems in circular cylindrical geometry.

A.9.6 KELVIN FUNCTIONS

When the Bessel functions have complex arguments, then the solution functions are known as 'Kelvin functions'. In this case the differential equation is of the form:

$$\frac{d^2y}{dx^2} + \frac{1}{x}\frac{dy}{dx} - \left(j + \frac{n^2}{x^2}\right) y = 0$$

Its solutions are:

$$y = \text{ber}_n(x) + j\ \text{bei}_n(x), \qquad \text{ber}_{-n}(x) + j\ \text{bei}_{-n}(x)$$
$$\text{ker}_n(x) + j\ \text{kei}_n(x), \qquad \text{ker}_{-n}(x) + j\ \text{kei}_{-n}(x)$$

where $\text{ber}_n(x) + j\,\text{bei}_n(x) = J_n\left\{x\exp\left(j3\frac{\pi}{4}\right)\right\} = \exp(jn\pi)\,J_n\left\{x\exp\left(-j\frac{\pi}{4}\right)\right\}$

$$= \exp\left\{j\frac{1}{2}n\pi\right\} I_n\left\{x\exp\left(j\frac{\pi}{4}\right)\right\} = \exp\left(j3n\frac{\pi}{4}\right) I_n\left\{x\exp\left(-j3\frac{\pi}{4}\right)\right\}$$

$$\text{ker}_n(x) + j\,\text{kei}_n(x) = \exp\left\{-j\frac{1}{2}n\pi\right\} K_n\left\{x\exp\left(j\frac{\pi}{4}\right)\right\}$$

$$= j\frac{\pi}{2} H_n^{(1)}\left\{x\exp\left(j3\frac{\pi}{4}\right)\right\} = -j\frac{\pi}{2} H_n^{(2)}\left\{x\exp\left(-j\frac{\pi}{4}\right)\right\}$$

When n is a +ve integer or zero,

$$\text{ber}_{-n}(x) = (-1)^n\,\text{ber}_n(x), \qquad \text{bei}_{-n}(x) = (-1)^n\,\text{bei}_n(x)$$
$$\text{ker}_{-n}(x) = (-1)^n\,\text{ker}_n(x), \qquad \text{kei}_{-n}(x) = (-1)^n\,\text{kei}_n(x)$$

When n is real, and x is real and non-negative

$$\text{ber}_{-n}(x) = \cos(n\pi)\,\text{ber}_n(x) + \sin(n\pi)\,\text{bei}_n(x) + \frac{2}{\pi}\sin(n\pi)\,\text{ker}_n(x)$$

$$\text{bei}_{-n}(x) = -\sin(n\pi)\,\text{ber}_n(x) + \cos(n\pi)\,\text{bei}_n(x) + \frac{2}{\pi}\sin(n\pi)\,\text{kei}_n(x)$$

$$\text{ker}_{-n}(x) = \cos(n\pi)\,\text{ker}_n(x) - \sin(n\pi)\,\text{kei}_n(x)$$

$$\text{kei}_{-n}(x) = \sin(n\pi)\,\text{ker}_n(x) + \cos(n\pi)\,\text{kei}_n(x)$$

A.9.7 LEGENDRE FUNCTIONS

By solving the three-dimensional Laplace's equation in spherical polar coordinate system by the method of separation of variables, the ODE in θ-variable is the Legendre equation (this has been derived in Section 4.2.7). So we write this equation directly as {i.e. Eq. (4.72)}:

$$(1-\mu^2)\frac{d^2\Theta}{d\mu^2} - 2\mu\frac{d\Theta}{d\mu} + \left\{n(n+1) - \frac{m^2}{1-\mu^2}\right\}\Theta = 0$$

where $\mu = \cos\theta$ and $\Theta = \Theta(\theta)$.

The two solutions of this second order differential equation are the "Associated Legendre Functions of the First and the Second Kinds. They are denoted as $\text{P}_n^m(\mu)$ and $\text{Q}_n^m(\mu)$ respectively. As with the Bessel functions, these functions are of degree n, order m and the argument is μ.

To solve the above equation, it can be simplified by a convenient substitution given below as

$$\Theta = (1-\mu^2)^{m/2}\,\Phi$$

whereby the differential equation becomes:

$$(1-\mu^2)\frac{d^2\Phi}{d\mu^2} - 2(m+1)\mu\frac{d\Phi}{d\mu} + (n-m)(n-m+1)\Phi = 0$$

which is very similar to the Legendre equation in two dimensions when there is no ϕ variation and hence $m = 0$. The associated Legendre equation then simplifies to the ordinary Legendre equation which is

$$(1-\mu^2)\frac{d^2\Theta}{d\mu^2} - 2\mu\frac{d\Theta}{d\mu} + n(n+1)\Theta = 0$$

Its solutions are the ordinary Legendre functions of the first and second kind respectively and are denoted by $P_n(\mu)$ and $Q_n(\mu)$ respectively, having degree n and argument μ. The values of $P_u(\mu)$ and $Q_u(\mu)$ are given in Eqs. (4.74) and (4.75) respectively.

The values of first few Legendre functions (or Legendre polynomials) are:

$$P_0(\mu) = 1$$

$$P_1(\mu) = \mu$$

$$P_2(\mu) = \frac{1}{2}(3\mu^2 - 1)$$

$$P_3(\mu) = \frac{1}{2}(5\mu^2 - 3\mu)$$

$$P_4(\mu) = \frac{1}{8}(35\mu^4 - 30\mu^2 + 3)$$

$$Q_0(\mu) = \tanh^{-1}\mu = \frac{1}{2}\ln\left(\frac{1+\mu}{1-\mu}\right)$$

$$Q_1(\mu) = \mu Q_0(\mu) - 1$$

$$Q_2(\mu) = P_2(\mu)\,Q_0(\mu) - \frac{3\mu}{2}$$

$$Q_3(\mu) = P_3(\mu)\,Q_0(\mu) - \frac{5\mu^2}{2} + \frac{2}{3}$$

$$Q_4(\mu) = P_4(\mu)\,Q_0(\mu) - \frac{35\mu^3}{8} + \frac{55\mu}{24}$$

The associated functions are:

$$P_n^m(\mu) = (1-\mu^2)^{m/2}\frac{d^m}{d\mu^m}\{P_n(\mu)\}$$

$$Q_n^m(\mu) = (1-\mu^2)^{m/2}\frac{d^m}{d\mu^m}\{Q_n(\mu)\}$$

for $|\mu| < 1$

A.9.8 ORTHOGONAL FUNCTIONS

Definition of Orthogonality: Given a set of functions, $\phi_o(x)$, $\phi_1(x)$, $\phi_2(x)$, ... which is denoted by $\{\phi_n(x)\}$. The set is said to be orthogonal in the interval (a, b) if

$$\int_a^b \phi_m(x) \cdot \phi_n(x)\, dx = 0$$

for any two distinct values of m and n, i.e. – $m \neq n$.

This orthogonality condition is used in the evaluation of coefficients of Fourier series expansions whereby all terms of the summation except one come out to be zero. Another point to be noted is that orthogonal functions would be oscillatory in nature. So the trigonometric functions, ordinary Bessel functions, Legendre functions are orthogonal, whereas hyperbolic functions and modified Bessel functions are not orthogonal.

Appendix 10

Bicylindrical Coordinate System

A.10.1 INTRODUCTION

It has been mentioned in Chapter 18, that the transmission lines, which are one way of guiding the electromagnetic waves, have been analysed by using the circuit concepts in terms of the potential drops across the lines and the current flowing through them (i.e. by using the terminal characteristics like the impedances and the admittances instead of the field vectors **E** and **H**). This approach made the analysis much simpler and it was also shown that the operating equation in terms of V and I comes out to be a one-dimensional wave equation (which indirectly proves the correctness of the equivalent circuit of the system). In fact, transmission lines consisting of two parallel circular cylindrical conductors (even of differing diameters) can be analysed as a field problem, but the coordinate system required for such analyses is the "Bicylindrical coordinate system", which consists of two families of orthogonal circles in the x–y plane. So we shall now briefly discuss the derivation of such a system.

A.10.2 Derivation of Bicylindrical Coordinate System

This coordinate system is generated by taking two families of orthogonal circles lying in the x–y plane (i.e. one family having their centres on the x-axis, and the other family (intersecting the circles of the first family orthogonally) having their centres on the y-axis), and translating these circles parallel to the z-axis, thereby forming sets of right circular cylinders.

Of the two families of circles, one family consists of two sets of non-intersecting circles, with all their centres collinear on the x-axis. The circles on the left side of the y-axis (i.e. $x < 0$) are the mirror image of the circles on the right side (i.e. $x > 0$), with the centres of the circles of zero-radius being located at $x = \pm a$ which are the two poles of the system (Figure A.10.1). The second family, which consists of circles orthogonal to the circles of the first family, has all the centres collinear on the y-axis. All the circles of this family pass through the poles $x = \pm a$, i.e. they all intersect at $x = \pm a$. A rigorous way of producing such circles (which are the sections of the coordinate surfaces) is by a suitable conformal transformation, which in this case is:

$$z^* = \frac{a(e^w + 1)}{e^w - 1} \tag{A.10.1}$$

where $z = x + jy, \quad w = u + jv;$ and

z^* = complex conjugate of $z = x - jy$

(Figure A.10.1 shows this transformation.)

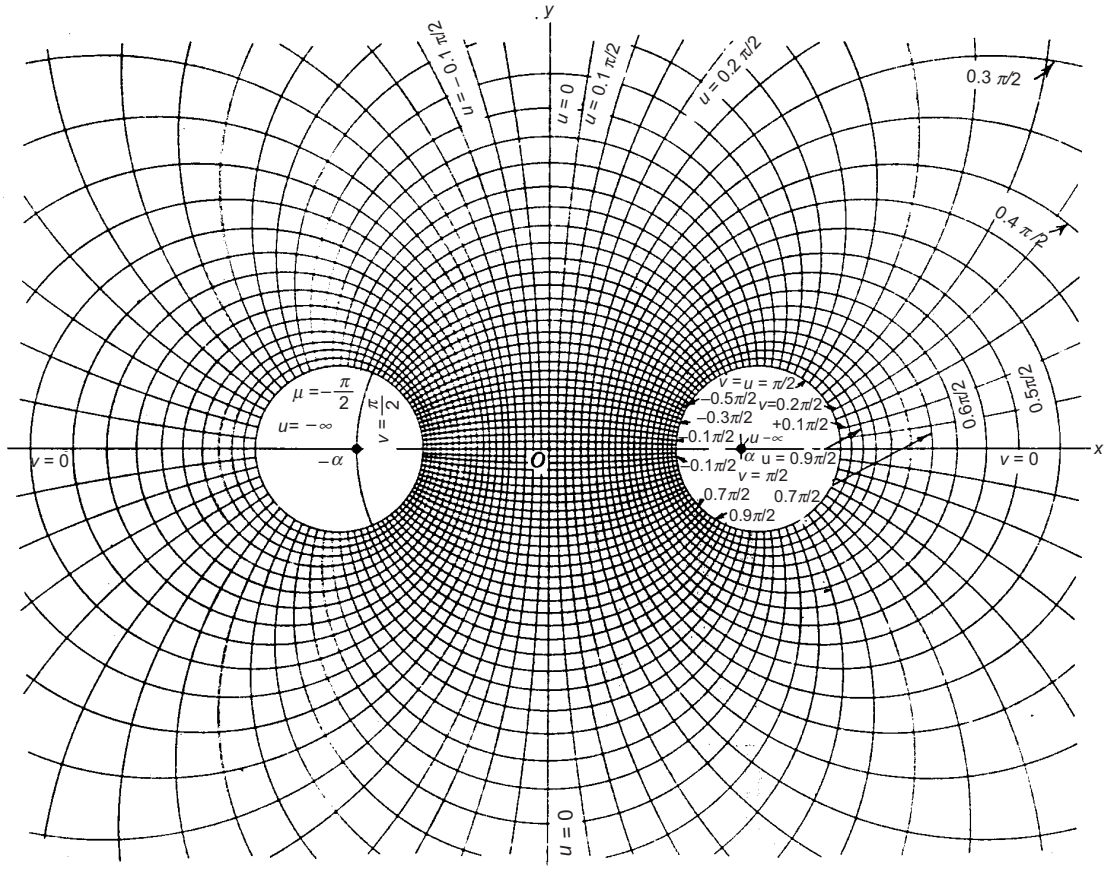


Figure A.10.1 Bipolar coordinates obtained by the transformation $z^* = \dfrac{a(e^w + 1)}{e^w - 1}$ and inversion of polar coordinates.

$$\therefore \qquad z^* = x - jy = \frac{a\{e^u(\cos v + j\sin v) + 1\}}{e^u(\cos v + j\sin v) - 1}$$

$$= \frac{[a\{e^u \cos v + 1\} + j\, a e^u \sin v][\{e^u \cos v - 1\} - j\, e^u \sin v]}{\{e^u \cos v - 1\}^2 + \{e^u \sin v\}^2}$$

$$\text{Numerator} = [a\{e^{2u}\cos^2 v - 1\} + a e^{2u}\sin^2 v] + j[ae^u \sin v\{e^u \cos u - 1\} - e^u \sin v\{e^u \cos v + 1\}a]$$

$$= [a\{e^{2u} - 1\} + j\{-2a e^u \sin v\}]$$

$$= 2a e^u\{\sinh u - j \sin v\}$$

$$\text{Denominator} = e^{2u}\cos^2 v - 2e^u \cos v + 1 + e^{2u}\sin^2 v$$

$$= (e^{2u} + 1) - 2e^u \cos v$$

$$= 2\,e^u (\cosh u - \cos v)$$

$\therefore$ Equating the real and the imaginary parts,

$$x = \frac{a \sinh u}{\cosh u - \cos v}, \qquad y = \frac{a \sin v}{\cosh u - \cos v} \tag{A.10.2}$$

Figure A.10.1 shows the two families of circles characterized by the above equations, where u = constant circles have their centres on the x-axis and the poles of the family are $x = \pm a$ (i.e. $u \to \pm \infty$ respectively).

So now we define a new coordinate system, which has been obtained from the rectangular coordinates in the z-plane, as the orthogonal circles system in the w plane and is given the new notation (η, θ, z) where u above has been replaced by η and v by θ. Thus, the circles in the bicylindrical coordinate system are characterized by:

$$x = \frac{a \sinh \eta}{\cosh \eta - \cos \theta}, \qquad y = \frac{a \sin \theta}{\cosh \eta - \cos \theta} \tag{A.10.3}$$

In the x–y plane, the circles η = constant are drawn about the two poles $x = \pm a$ and the orthogonal circles of the other family (i.e. θ = constant) have their centres on the y-axis and they all pass through the poles $x = \pm a$.

The cylindrical coordinate system is generated by translating all the circles parallel to the z-axis, so that the two orthogonal families of right circular cylinders are produced. The three mutually perpendicular coordinate surfaces are: the cylinders η = constant, the cylinders θ = constant, and the planes z = constant. The new coordinates (η, θ, z) are called the bicylindrical coordinates and are designated by the equations (which relate them to the Cartesian coordinates):

$$x = \frac{a \sinh \eta}{\cosh \eta - \cos \theta}, \qquad y = \frac{a \sin \theta}{\cosh \eta - \cos \theta}, \qquad z = z \tag{A.10.4}$$

Figure A.10.2 shows the details of this coordinate system.

A.10.2.1 Some Geometrical Aspects (Figure A.10.2)

The equation of any circle of radius r, not necessarily a member of the coordinate surface η = constant, with its centre on the x-axis at the point $x = w$ is:

$$(x - w)^2 + y^2 = r^2 \tag{A.10.5}$$

For this circle to be a member of the coordinate surfaces, i.e. one of the circles η = constant, we substitute for x and y from the equations (A.10.3) in the above equation and get:

$$\left(\frac{a \sinh \eta}{\cosh \eta - \cos \theta} - w\right)^2 + \frac{a^2 \sin^2 \theta}{(\cosh \eta - \cos \theta)^2} = r^2$$

This gives

$$w = a \coth \eta, \qquad r = \frac{a}{|\sinh \eta|} \tag{A.10.6}$$

where, of course, a = semi-polar distance (on the x-axis) for the specified coordinate system.

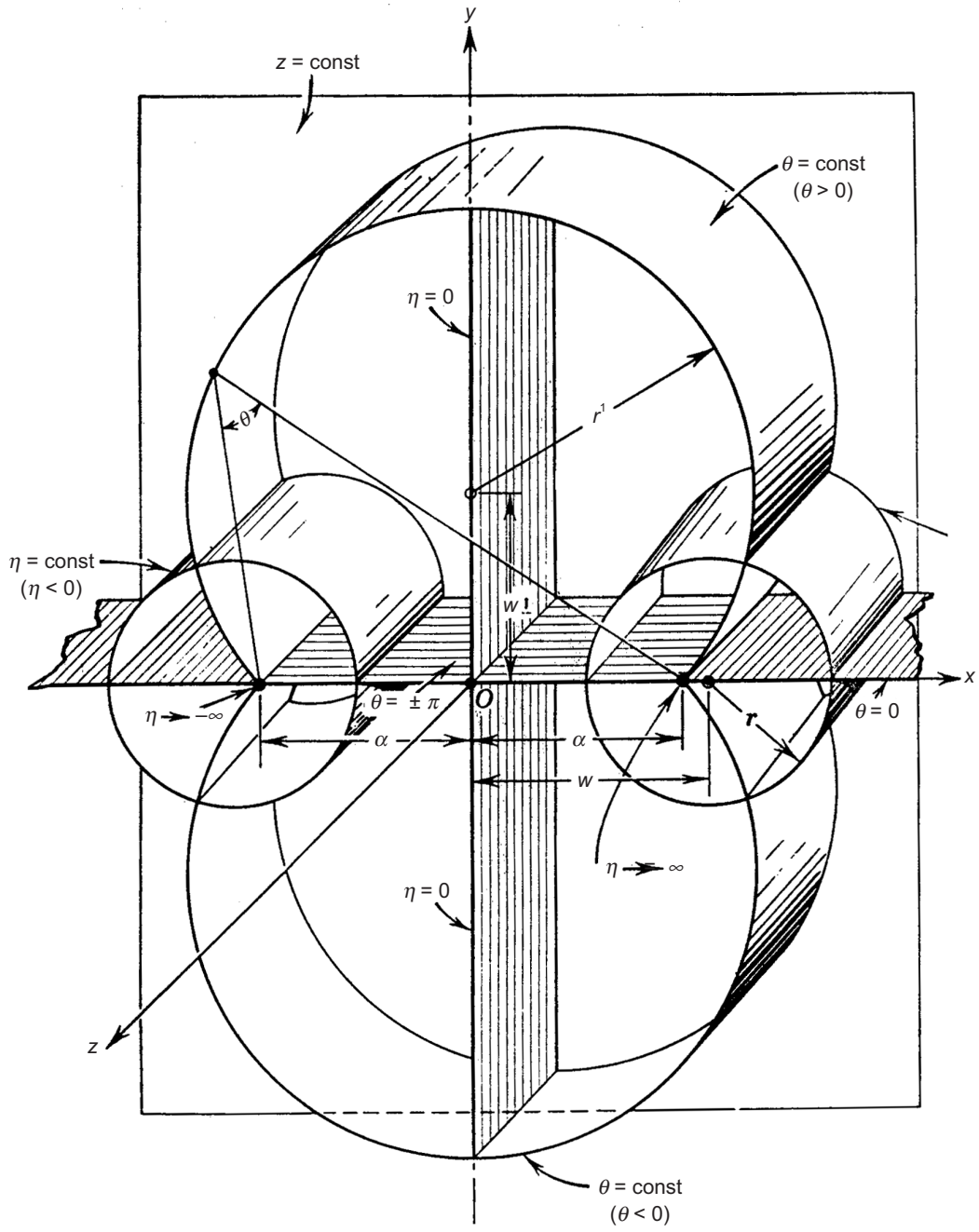


Figure A.10.2 Bicylindrical coordinates. The surfaces η = const are circular cylinders with axes in the *xz*-plane, surfaces θ = const are portions of circular cylinders with axes in the *yz*-plane, surfaces z = const are parallel planes.

Thus, both the position of the centre (w, 0) the radius (= r) of the circle depend on the polar distance (= $2a$) and on η. So, as $\eta \to 0$ the radius $r \to \infty$, and as $\eta \to \infty$ the radius $r \to 0$. Hence the centres of the η = constant circles get translated along the x-axis as η changes; and when the

radius r becomes zero (i.e. $r \to 0$) the centre of the zero-radius circle gets located at the pole $x = a$ for which $\eta \to \infty$. On the other hand, as $x \to \infty$, $\eta \to 0$.

So, it is seen that since these three geometrical parameters a, w and r are interrelated, it is possible to draw η = constant circles with their centres (w, 0) located relative to the position of the pole, and also obtain the radius of the circle. The relationship between these three quantities can be expressed as:

$$w^2 = a^2 \coth^2 \eta = a^2 \left(\frac{1 + \sinh^2 \eta}{\sinh^2 \eta} \right) = r^2 + a^2 \tag{A.10.7}$$

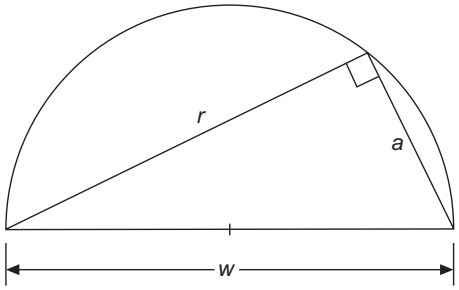


Figure A.10.3 Relationship between *w*, *a* and *r* (for η = constant circles).

Thus, if for a given η, its centre is specified, then the pole and the radius can be found; or if the pole is specified, then the centre of the circle and its radius can be obtained; or if the radius of the circle is specified, then the pole and the centre of the circle can be located. Usually the polar distance $2a$ is specified for a given coordinate system. Another point to be noted about the η circles is that since for $\eta = 0$, the radius r is infinite, the yz-plane (i.e. $x = 0$) is the $\eta = 0$ surface and the centre of this circle ($\eta \to 0$) also tends to infinity on the x-axis. A physical basis for the construction of these circles will be discussed later.

Next, for θ = constant circles, any circle (not necessarily θ-circle only) with its centre on the y-axis is given by

$$x^2 + (y - w')^2 = r'^2 \tag{A.10.8}$$

For this circle to be a member of the coordinate surfaces θ-const., we substitute for x and y from the equations (A.10.3) (and it can be checked that)

$$w' = a \cot \theta \qquad \text{and} \qquad r' = \frac{a}{|\sin \theta|} \tag{A.10.9}$$

All these circles (i.e. sections of θ = constant cylinders) pass through the two poles $x = \pm a$ and intersect the (η = constant) circles orthogonally. (All the angular and linear relationships are shown in Figure A.10.4.)

Thus, w' = distance of the centre of a (θ = constant) circle from the origin of the co-ordinate system = $a \coth \theta$;

and r = the radius of a (θ = constant) circle $= \dfrac{a}{|\sin \theta|}$.

It is obvious from the above figure that the parameter θ represents the angle subtended by the bi-polar distance AA' (= $2a$) on to the part of the circle (θ = constant) on the +ve side of the

y-axis. Hence, as the angle $AOA' = \pi$, the strip of width AA' (= $2a$) in the xz-plane would represent $\theta = \pi$.

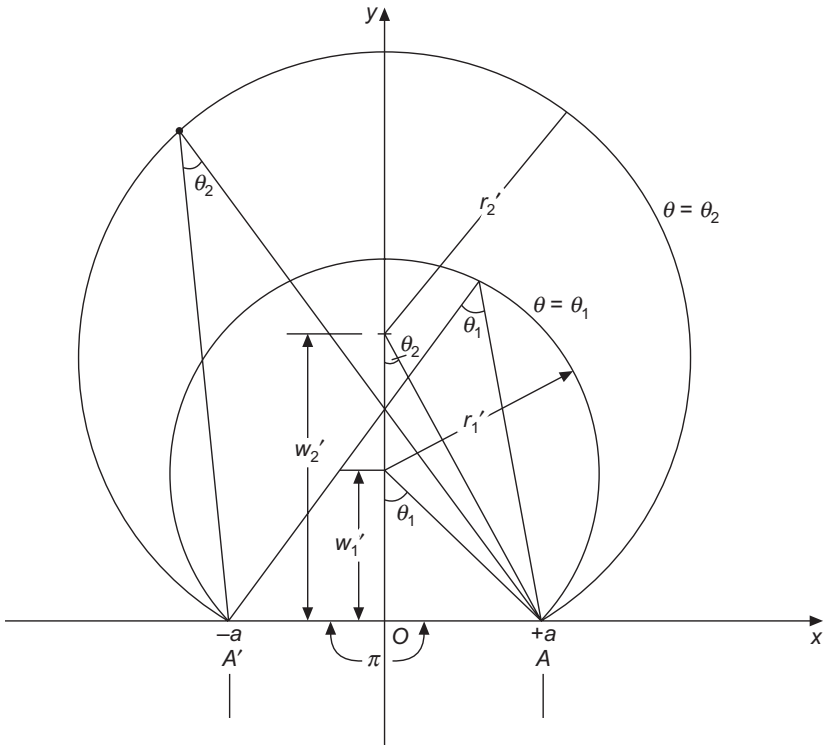


Figure A.10.4 Relationship for (θ = constant) circles.

The rest of the xz-plane (i.e. from $x = a$ to $x \to +\infty$ and $x = -a$ to $x \to -\infty$) would then represent $\theta = 0$.

The portions of the cylinders (θ = constant) above the xz-plane, for positive values of y, are designated by the positive values of θ, and those below the xz-plane are designated by the negative values of θ.

Thus (η = constants) represent a system of co-axial circles of the non-intersecting type, having limiting points at $x = \pm a$, and the radical axis is $y = 0$ (i.e. $x = \pm a$ are the two poles, and x-axis is the common axis on which the centres of the system of these co-axial circles lie with the poles ($x = \pm a$) being the centres of the zero-radius circles).

The (θ = constants) circles are another system of co-axial circles of the intersecting type having $x = \pm a$ as common points for all the members of this set. The radical axis (i.e. the line on which the centres of the these circles lie) for this system is $x = 0$.

These two systems of co-axial circles are orthogonal.

A.10.3 SOLUTION OF THE FIELD IN THE BICYLINDRICAL COORDINATE SYSTEM

The metric coefficients (i.e. g_{ir} and g_{ij}) are obtained from equations (A.10.4) as

$$g_{11} = g_{22} = \frac{a^2}{(\cosh \eta - \cos \theta)^2} = \sqrt{g}; \quad g_{33} = 1 \tag{A.10.10}$$

It should be noted that the Laplace's equation is not separable in this coordinate system, but in two-dimensions, if the function (say ϕ) is independent of z, then the Laplace's equation simplifies to:

$$\frac{\partial^2 \phi}{\partial \eta^2} + \frac{\partial^2 \phi}{\partial \theta^2} = 0 \tag{A.10.11}$$

whose solutions are of the form:

$$\phi = e^{\pm p\theta} \begin{pmatrix} \sin \\ \cos \end{pmatrix} p\eta$$

$$\phi = e^{\pm p\eta} \begin{pmatrix} \sin \\ \cos \end{pmatrix} p\theta \tag{A.10.12}$$

When ϕ is a function of one variable only, the operating equation reduces to:

$$\frac{d^2 \phi}{d\eta^2} = 0, \qquad \phi = A + B\eta$$

or

$$\frac{d^2 \phi}{d\theta^2} = 0, \qquad \phi = A + B\theta \tag{A.10.13}$$

A.10.4 AN ALTERNATIVE PHYSICAL BASIS FOR THE DERIVATION OF (η = CONSTANT) CIRCLES IN THE COORDINATE SYSTEM

So far we have derived this coordinate system on a rigorous basis purely from mathematical considerations (i.e. conformal transformation and geometrical concepts). However it is possible to derive this system starting from a physical problem as well. We shall now discuss such an approach briefly.

We consider a system of two parallel wires, each of circular cross-section of radius r_0, the distance between their centres being $2w$ as shown in Figure A.10.5.

The two wires carry surfaces charges $\pm \rho_L$ per unit length. These two wires can be replaced by two line charges $+\rho_L$ and $-\rho_L$ (per unit length). The exact location of these two line charges can be found from the condition that the surfaces of the two metal wires (of radius r_0) remain equipotential surfaces. Also all the other cylindrical equipotential surfaces, surrounding each of these wires (though not concentrically), remain unchanged. These requirements imply that the tangential electric fields on these surfaces will always be equal to zero. (The same results would be obtained if instead of treating this problem as that of electrostatics, we considered the conductors to carry currents $\pm I$ and applied the required conditions to the magnetic field of the currents).

The potential $V(x, y)$ at the point P (in Figure A.10.5) is given by

$$V(x, y) = \frac{\rho_L}{2\pi\varepsilon} \ln \frac{a}{r_2} + \frac{-\rho_L}{2\pi\varepsilon} \ln \frac{a}{r_1} = \frac{\rho_L}{2\pi\varepsilon} \ln \frac{r_1}{r_2} \tag{A.10.14}$$

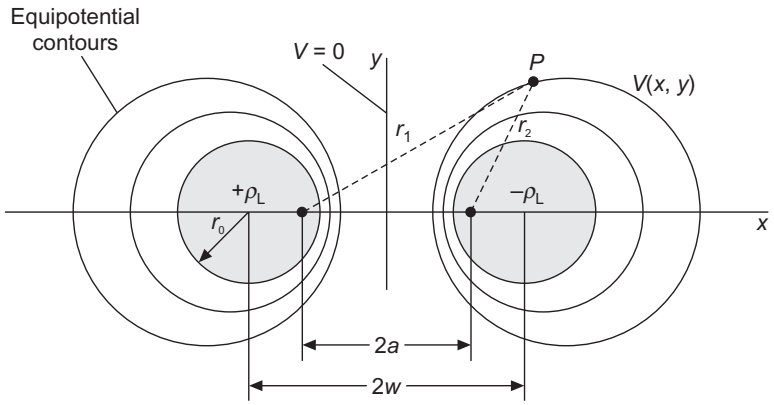


Figure A.10.5 Equipotential contours surrounding two line charges $+\rho_L$ and $-\rho_L$. The surfaces at $r = a$ are equipotential surfaces, therefore the electric field will always be normal to the metal surfaces.

where r_1 and r_2 are the distances of the point P from the locations of the equivalent line charges $\pm \rho_L$ at the points $x = \pm a$ as shown. The x-axis of the used coordinate system is the line passing through the centres of the conductors, and the y-axis is passing through the mid-point of the section joining the centres. So the y-axis (or yz-plane) would be also an equipotential surface at zero potential. All other equipotential contours can be obtained by applying the condition that

$$\frac{r_1}{r_2} = \frac{\sqrt{\{(a+x)^2 + y^2\}}}{\sqrt{\{(a+x)^2 + y^2\}}} = k \tag{A.10.15}$$

where k is any arbitrary constant (at this stage).

This can be re-written as:

$$(a + x)^2 + y^2 = k^2 \{(a - x)^2 + y^2\}$$

or (after a bit of algebraic manipulations)

$$\left(x - a\frac{k^2+1}{k^2-1}\right)^2 + y^2 = \left(\frac{2ka}{k^2-1}\right)^2 \tag{A.10.16}$$

This form is obtained by adding the term $a^2\left(\frac{k^2+1}{k^2-1}\right)$ to both sides of the equation.

Thus, Eq. (A.10.16) is the equation to a circle whose centre lies on the x-axis at the point (w, 0), where

$$w = a\frac{k^2+1}{k^2-1}$$

and whose radius is r_0 where (A.10.17)

$$r_0 = \frac{2ka}{k^2-1}$$

Since k is any arbitrary constant, we can put

$$k = e^{\eta}$$

(remembering that η is constant for such circles), and then

$$w = a\frac{k^2+1}{k^2-1} = a\frac{k+\dfrac{1}{k}}{k-\dfrac{1}{k}} = a\frac{e^{\eta}+e^{-\eta}}{e^{\eta}-e^{-\eta}}$$
$$= a\frac{\cosh\eta}{\sinh\eta} = a\coth\eta \qquad \text{(A.10.18)}$$

$$r_0 = \frac{2ka}{k^2-1} = \frac{2a}{k-\dfrac{1}{k}} = \frac{2a}{e^{\eta}-e^{-\eta}}$$
$$= \frac{a}{\sinh\eta} \qquad \text{(A.10.19)}$$

which are the same, as derived in Section A.10.2.

So, as before

$$w^2 = a^2\coth^2\eta = \frac{a^2(1+\sinh^2\eta)}{\sinh^2\eta} = r_0^2 + a^2$$

Thus if the radius of the two wires and the distance between their centres (i.e. r_0 and w respectively) are known, then the position of the equivalent line charges (i.e. $x \pm a$) can be determined as shown above. The (η = constant) circles are the equipotential lines, and the orthogonal (θ = constant) circles, passing through the two poles at $x = \pm a$ are the lines of force (or flow lines or flux lines) and they can be determined as before.

The use of this coordinate system has been demonstrated in a number of problems (i.e. Problems 2.48, 12.52, 12.53) in the companion volume *Electromagnetism: Problems and Solutions*, 2nd Edition, 2008, ranging from low frequency power engineering to high frequency transmission lines.

Appendix 11

Method of Moments

A.11.1 INTRODUCTION

The method of moments (MoM) was first used by R.F. Harrington who called it 'moment method'. The method is conceptually simple and is commonly used in solving integral equations in regions which are open. It has wide applications in electromagnetics, particularly for designing and analysing problems pertaining to antennae, problems of scattering, transmission-line capacitance calculations as well as metallic object capacitances, interpretation of high frequency signals from aircraft during flight. In electrostatics, this method is used (for example) for determining the unknown charge distribution on conductors when their potentials are specified. It is thus seen that this method has been and is being used for integral solutions of Poisson's equation and wave equation.

So we start with the general definition of "integral equation" which involves an unknown function in its integrand and has the general form

$$f(x) = \int_a^b K(x, t)\ \phi(t)\ dt \tag{A.11.1}$$

where the functions $f(x)$, $K(x, t)$ and the integral limits a, b are known quantities. The quantity to be determined is the unknown function $\phi(t)$, and the known function $K(x, t)$ is called the 'kernel' of the equation. MoM is a common numerical technique to solve such equations.

Before we discuss the general steps of the method, we will explain it by solving a simple problem by MoM so as to make it easier to understand. So we choose a Poissonian field problem in electrostatics. Generally, if the charge distribution ρ is known, then the potential V, the electric field **E** and the total charge **Q** can be evaluated in a straightforward manner. The converse problem is: finding the charge distribution ρ, when the potential V is known. The integral expressing the relationship between ρ and V is given by

$$V = \iiint \frac{\rho dv}{4\pi\varepsilon r} \tag{A.11.2}$$

However to make the problem easier for our understanding we will replace the triple integral of the three-dimensional problem by a one-dimensional one. Instead of volume charge density ρ, we shall consider a line charge density ρ_L on a thin conducting wire of length L, i.e.

$$V_0 = \int_0^L \frac{\rho_L dl}{4\pi\varepsilon_0 r} \tag{A.11.3}$$

We position the wire along the x-axis, keeping one end at the origin and since the above equation applies to any observation point along the x-axis, we choose an observation point x_k (which is known as the match point). (see Figure A.11.1) and the integral can then be expressed as

$$V_0 = \frac{1}{4\pi\varepsilon_0} \int_0^L \frac{\rho_L(x)dx}{|x_k - x|} \tag{A.11.4}$$

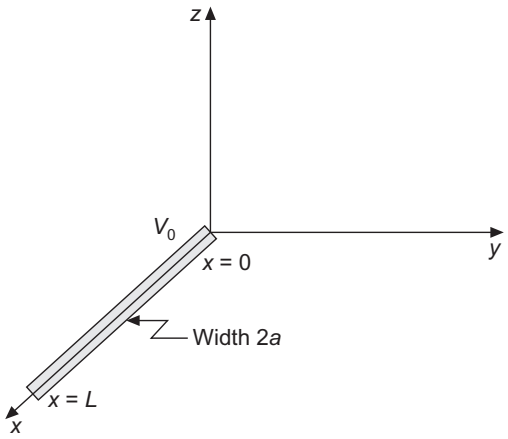


Figure A.11.1 Conducting wire maintained at constant potential V_0.

Since integration is finding the area under c curve, then in the present problem,

$$\int f(x)\,dx = f(x_1)\,\Delta x + f(x_2)\,\Delta x + \cdots + f(x_N)\,\Delta x$$

$$= \sum_{k=1}^{N} f(x_k)\,\Delta x \tag{A.11.5}$$

Δx being a small elemental length along the x-axis. Here the length L has been sub-divided in N elements each of length Δ. Substituting from Eq. (A.11.5) into Eq. (A.11.4), we get

$$4\pi\varepsilon_0 V_0 = \frac{\rho\Delta}{|x_k - x_1|} + \frac{\rho\Delta}{|x_k - x_2|} + \cdots + \frac{\rho\Delta}{|x_k - x_N|} \tag{A.11.6}$$

where $$\Delta = \frac{L}{N} = \Delta x \quad \text{(Figure A.11.2)}$$

The implicit assumption in the above equation is that unknown charge density on each segment of the wire is constant.

Since Eq. (A.11.6) holds for any point x_k on the wire, by giving x_k all values from x_1 to x_N, we get N similar equations, i.e.

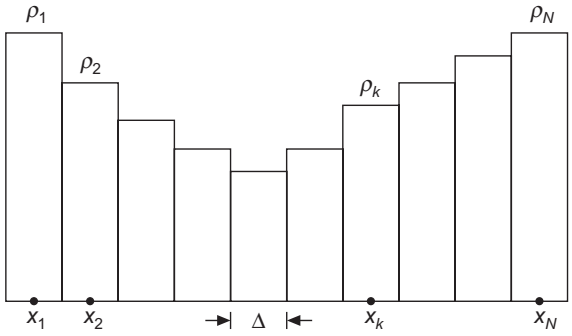


Figure A.11.2 Charged wire sub-divided into N segments.

$$4\pi\varepsilon_0 V_0 = \frac{\rho_1\Delta}{|x_1 - x_1|} + \frac{\rho_2\Delta}{|x_1 - x_2|} + \cdots + \frac{\rho_N\Delta}{|x_1 - x_N|} \tag{A.11.7a}$$

$$4\pi\varepsilon_0 V_0 = \frac{\rho_1\Delta}{|x_2 - x_1|} + \frac{\rho_2\Delta}{|x_2 - x_2|} + \cdots + \frac{\rho_N\Delta}{|x_2 - x_N|} \tag{A.11.7b}$$

$$4\pi\varepsilon_0 V_0 = \frac{\rho_1\Delta}{|x_N - x_1|} + \frac{\rho_2\Delta}{|x_N - x_2|} + \cdots + \frac{\rho_N\Delta}{|x_N - x_N|} \tag{A.11.7c}$$

The idea of matching the LHS of Eq. (A.11.4) with the RHS (match points) of the above equation is equivalent to taking moments in mechanics (and this is the reason for the name of the method). If the observation points are located at the centres of the segments, then the above set of equations can be expressed in matrix form as

$$[B] = [A]\ [\rho] \tag{A.11.8}$$

where

$$[B] = 4\pi\varepsilon_0 V_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \qquad A = \begin{bmatrix} A_{11} & A_{12} & A_{1N} \\ A_{21} & A_{22} & A_{2N} \\ A_{N1} & A_{N2} & A_{NN} \end{bmatrix} \tag{A.11.9a}$$

$$\text{where} \quad A_{mn} = \frac{\Delta}{|x_m - x_n|} \qquad (m \neq n)$$

$$\text{and} \qquad [\rho] = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_N \end{bmatrix} \tag{A.11.9b}$$

$[\rho]$ is the matrix with unknown elements, and it can be determined by either matrix inversion or Gaussian elimination.

i.e. $$[\rho] = [A]^{-1}\,[B] \tag{A.11.10}$$

Care must be taken in evaluating A_{nn}—the diagonal terms. Since the conducting wire carries charge, the observation points have been chosen at the centre of the segment so that the limits of integration in x-direction will be $x = \pm\dfrac{\Delta}{2}$. To evaluate A_{nn},

$$V_{\text{centre}} = \frac{1}{4\pi\varepsilon_0}\int_0^{2\pi}\int_{-\Delta/2}^{+\Delta/2}\frac{\rho_S a\,d\phi\,dx}{(a^2 + x^2)^{1/2}}$$

$$= \frac{2\pi a\rho_S}{4\pi\varepsilon_0}\ln\left[\frac{\Delta/2 + \{(\Delta/2)^2 + a^2\}^{1/2}}{-\Delta/2 + \{(\Delta/2)^2 + a^2\}^{1/2}}\right]$$

where ρ_S is the surface charge density on the conductor.

Assuming $\Delta >> a$,

$$V_{\text{centre}} = \frac{2\pi a\rho_S}{4\pi\varepsilon_0}\,2\ln\left(\frac{\Delta}{a}\right) = \frac{2\rho_L}{4\pi\varepsilon_0}\ln\left(\frac{\Delta}{a}\right), \quad \rho_L = 2\pi a\rho_S \tag{A.11.10a}$$

$\therefore$ $$a_{nn} = 2\ln\left(\frac{\Delta}{a}\right) \tag{A.11.10b}$$

A.11.2 GENERAL CONCEPTS

Now that an actual problem has been solved, it will be easier to appreciate the general approach to the method.

Expressing the problem as an operator equation which has to be solved, we have

$$L\,f = g \tag{A.11.11}$$

where

L = linear operator (i.e. integro-differential operator)
f = the unknown function (i.e. field or induced current)
g = known function (i.e. the excitation).

The first step to determine f is to approximate it by a linear combination of functions, i.e. f_n, $n = 1, 2, ..., N$ which are called **'Expansion'** or **'Basis'** functions.

$$f \approx f_N^a = \sum_{n=1}^{N}\alpha_n f_n \tag{A.11.12}$$

where (α_n) are constants, and the Σ term is the N term approximation to f. [Refer to Eq. (A.11.5) of the Poissonian field problem solved by MoM.]

The next step is to substitute for f in the operator Eq. (A.11.11), from Eq. (A.11.12), and we obtain:

$$L\,f_N^a = \sum_{n=1}^{N}\alpha_n L\,f_n \approx g \tag{A.11.13}$$

The substitution for f is an approximate one which does not make the LHS exactly equal to the RHS (i.e. = g) of this equation. The difference between the two is called the "Residual" (= R_N), which is

$$R_N = g - Lf_N^a = g - \sum_{n=1}^{N} \alpha_n L f_n \tag{A.11.14}$$

Making the residual orthogonal to the set of functions (= W_m) called the weighting functions, we get:

$$\sum_{n=1}^{N} \alpha_n (W_m, Lf_n) = (W_m, g), \quad m = 1, 2, 3, ..., M \tag{A.11.15}$$

where we define

$$\text{Inner product: } \langle f, h \rangle = \int_D f^*(x)\, h(x)\, dx \tag{A.11.16a}$$

and for real f,

$$\text{Inner product: } \langle f, h \rangle = \int_D f(x)\, h(x)\, dx \tag{A.11.16b}$$

Now, Eq. (A.11.14) can be written in matrix form as

$$A\alpha = g \tag{A.11.17}$$

{Ref.: Compare Eqs. (A.11.14) and (A.11.15) with Eq. (A.11.7), and Eq. (A.11.8) with Eq. (A.11.17)}

where

$$A = \begin{bmatrix} A_{11} & A_{12} & ... & A_{1N} \\ A_{21} & A_{22} & ... & A_{2N} \\ ... & ... & ... & ... \\ A_{M1} & A_{M2} & ... & A_{MN} \end{bmatrix} = \text{Moment matrix}$$

$$g = [g_1 \quad g_2 \quad ... \quad g_M]^T \leftarrow \text{excitation vector}$$

and

$$\alpha = [\alpha_1 \quad \alpha_2 \quad ... \quad \alpha_N]^T$$

$$A_{mn} \approx \langle W_m, Lf_n \rangle \quad \text{and} \quad g_m \approx \langle W_m, g \rangle$$

This can be solved provided the expansion function is known.

So now we discuss briefly some of the types and the properties of different expansion functions and then weighting functions.

The requirements of any expansion function are:

1. It should be few and simple
2. It should approximate well with f.

There are two kinds of expansion functions, which are

(a) Sub-domain type
(b) Entire domain type

A. Sub-domain Expansion Functions

We consider first some of the sub-domain types of expansion functions.

1. δ-function Expansion Function

The unknown function f can be approximated by placing a number of Dirac δ-functions at discrete points along f. Hence it is applicable to (say) wire-grid model of electromagnetic surfaces. When the δ-function is being used as the expansion function, the method of Green's function is an alternative method and may be a preferable choice because of its greater generalization.

2. Pulse Function Expansion Function

This is the staircase approximation and mathematically the pulse function is:

$$P_n(x) = \begin{cases} 1, & x \in \Delta x_n; \\ 0, & \text{otherwise} \end{cases}$$

This function is quite popular as the integration is quite simple, but it can be used only if the operator L does not have derivatives (because the derivative of a pulse is an impulse).

3. Triangular Function

This is a piece-wise linear approximation and is more accurate than the pulse functions. Another advantage of this approximation is that, it is effective even when the operator L contains derivative terms, i.e. $\frac{\partial f}{\partial x}$. But it is unsuitable for problems which contain points where $f \to \infty$.

The mathematical representation of triangular functions with equal as well as unequal segment lengths are stated herewith:

(a) Δ^{lar} function with equal segment length

$$T_1(x) = \begin{cases} 1 - \dfrac{x - x_1}{\Delta}, & x_1 \le x \le x_2 \\ 0, & \text{otherwise} \end{cases}$$

$$T_n(x) = \begin{cases} 1 - \dfrac{|x_n - x|}{\Delta}, & |x - x_n| \le \Delta, \quad n = 1, 2, 3, \ldots, N \\ 0, & \text{otherwise} \end{cases}$$

$$T_N(x) = \begin{cases} 1 - \dfrac{x_N - x}{\Delta}, & x_{N-1} \le x \le x_N \\ 0, & \text{otherwise} \end{cases}$$

(b) Δ^{lar} function with unequal segments

$$T_1(x) = \begin{cases} 1 - \dfrac{x - x_1}{\Delta_1}, & x_1 \le x \le x_2 \\ 0, & \text{otherwise} \end{cases}$$

$$T_n(x) = \begin{cases} 1 - \dfrac{x_n - x}{\Delta_{n-1}}, & x_{n-1} \le x \le x_n \\ 1 - \dfrac{x - x_n}{\Delta_n}, & x_n \le x \le x_{n+1} \quad n = 2, 3, \ldots, N-1 \\ 0, \text{ otherwise} \end{cases}$$

$$T_N(x) = \begin{cases} 1 - \dfrac{x_N - x}{\Delta_{N-1}}, & x_{N-1} \le x \le x_N \\ 0, \text{ otherwise} \end{cases}$$

where $\Delta_n = x_{n+1} - x_n$ for $n = 1, 2, 3, \ldots, N-1$.

4. Piece-wise Simusoidal Function

This is a higher-order fit. (There can be quadratic and other higher-order functions as well.) Here sinusoidal function has been chosen. It should be noted that the period of each sinusoidal function is one wavelength. This function is defined as

$$f_n(x) = \begin{cases} \dfrac{\sin k\left(\Delta - |x - x_n|\right)}{\sin(k\Delta)}, & |x - x_n| \le \Delta \\ 0, \text{ otherwise} \end{cases}$$

where 2Δ = half wavelength

This function is normally used in wire antennae.

B. Entire Domain Expansion Functions

1. Trigonometric Functions, e.g.

$$F_n(x) = \begin{cases} A_n + B_n \sin\{k(x - x_n)\} + C_n \cos\{k(x - x_n)\}, & |x - x_n| \le \Delta \\ 0, \text{ otherwise} \end{cases}$$

2. Chebyshev Polynomials and Other Polynomials

These are:

Chebyshev polynomial of the First kind:

$T_n(\cos\theta) = \cos n\theta, \quad n = 0, 1, 2, \ldots$

Chebyshev polynomial of the second kind

$$U_n(\cos\theta) = \frac{\sin(n+1)\theta}{\sin\theta}, \quad n = 0, 1, 2, \ldots$$

These are used for wire antennae, strip lines, slot lines, etc.

3. Wave Functions

The solution of the Helmholtz equation is

$$f_n = Z_n(kr)e^{jn\psi}$$

where

Z_n = cylindrical Bessel functions for two-dimension problems

and = spherical Bessel functions or Legendre functions in three-dimensions

Next we consider weighting functions.

Weighting Functions

As in the case of expansion functions, there exists a wide choice of weighting functions. We will consider the following methods.

1. Point Matching Method

This is used when the boundary conditions are imposed only at finite isolated points. In this case, the weighting functions are "Dirac Delta functions".

$$W_m = \delta(x - x_m), \quad m = 1, 2, 3, ..., N$$

This method is simple to use. The accuracy of the method improves with closely spaced points which are spaced non-uniformly. This method is also known as "Collocation method". However, care must be taken while using the technique as the solution (of the Point matching method) does not always converge to the true solution.

2. Galerkin's Method

In this case, the weighting function is same as the expansion function.

The main advantage of this method is that a small number of functions is sufficient to produce an accurate result.

3. Least Square Method

In this case, the norm of the residual is minimized.

$$\|R_N\|^2 = \|g - \sum_{n=1}^{N} \alpha_n L f_n\|^2$$

where the norm of f is

$$\|f\| = (f \cdot f)^{1/2}$$

At the minimum, $\frac{\partial}{\partial \alpha_n} \|R_N\|^2 = 0$, which gives

$$\sum_{n=1}^{N} \alpha_n \langle Lf_m, Lf_n \rangle = \langle Lf_m, g \rangle, \quad m = 1, 2, ..., N$$

The moment matrix A has elements

$$A_{mn} = \langle Lf_m, Lf_n \rangle$$

Note: 1. A is positive definite Hermitian matrix.

2. LS method requires longer computing time than the point matching.

Plates

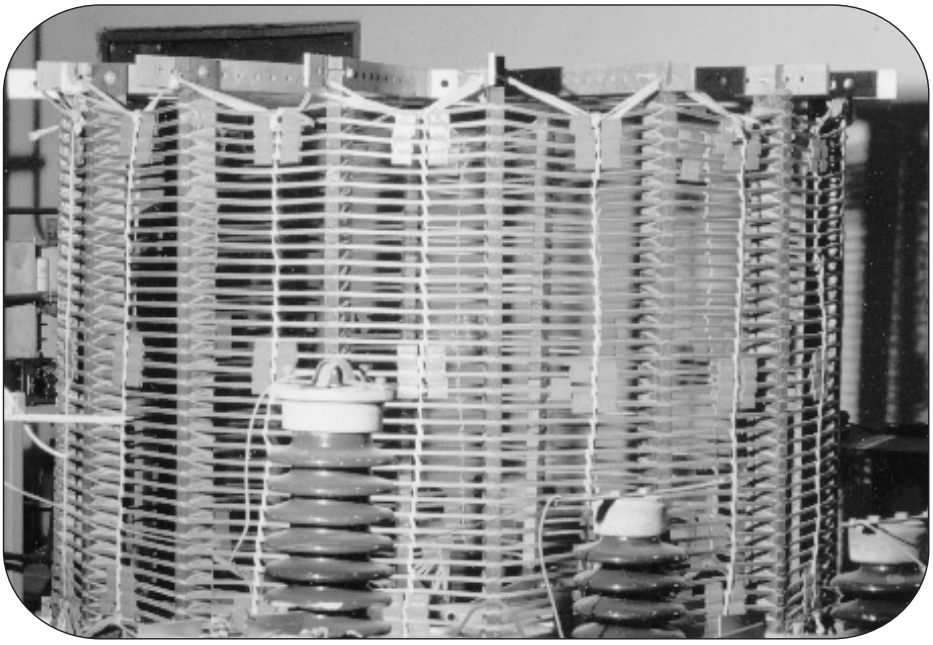


Plate 1 A view of the reactor subjected to high current short-circuit test.

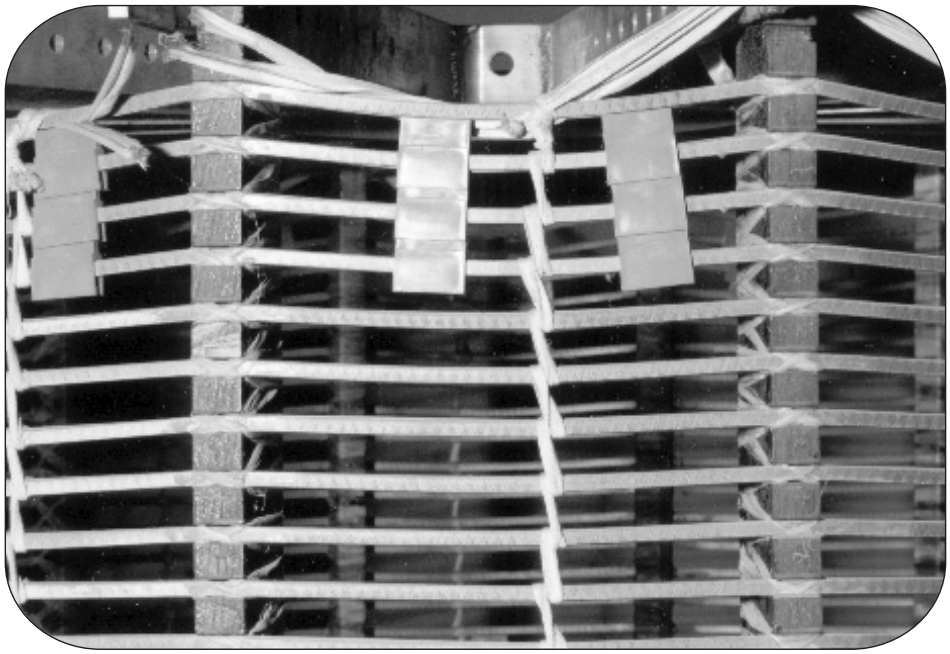


Plate 2 Another view of a section of the reactor shown in Plate 1.

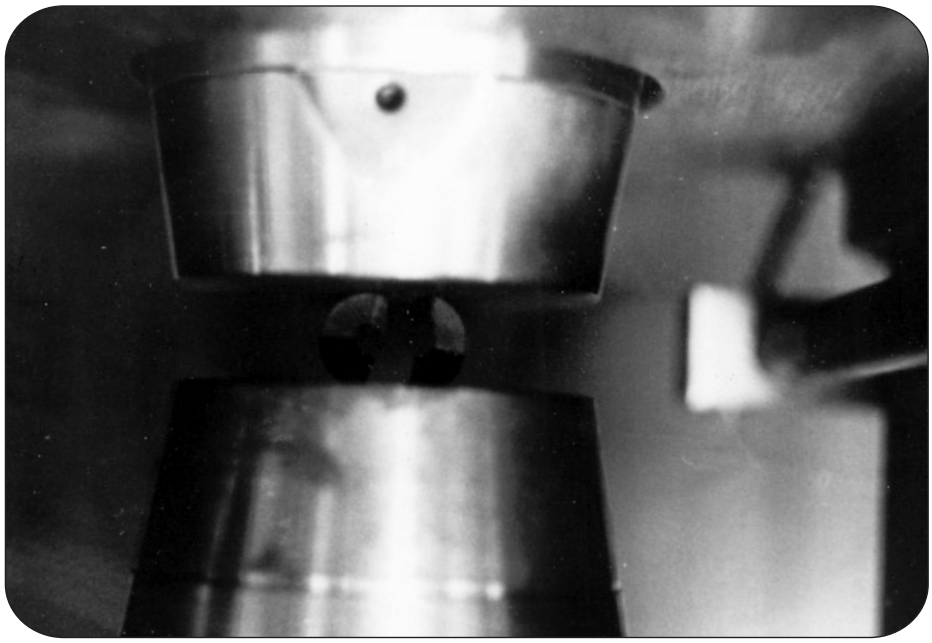


Plate 3 Effect of radial magnetization on an alnico ring.

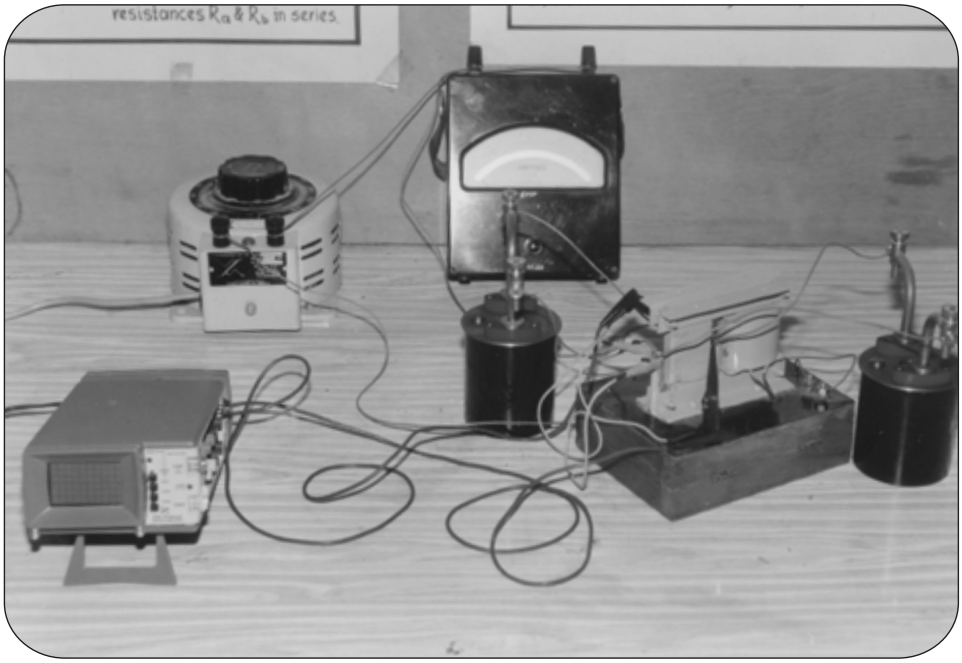


Plate 4 A view of the model experiment to demonstrate the double-valued P.D. in the circulating current circuit.

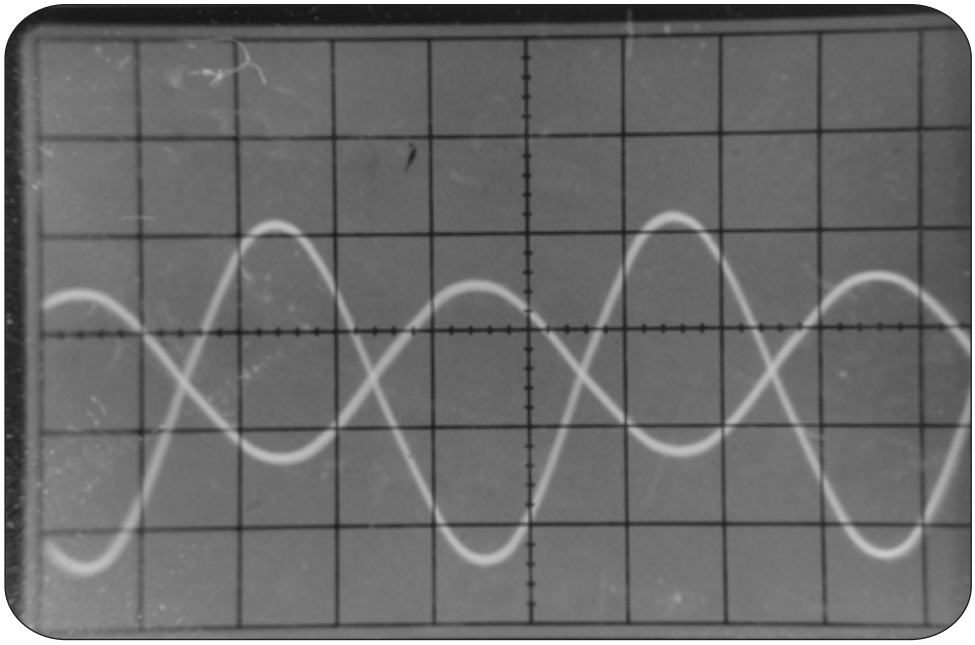


Plate 5 Wave shapes of V_a and V_b from the demonstration experiment (Plate 4) as observed on the CRO screen.